

PRE-TEST ESTIMATION IN A REGRESSION
MODEL WITH A MIS-SPECIFIED
ERROR COVARIANCE MATRIX

A thesis presented for the
degree of Doctor of Philosophy
in Economics
at the
University of Canterbury
Christchurch
New Zealand

by

K.V. Albertson

January 1993

HA
31.7
A334
1993

Acknowledgements

I am deeply indebted to Professor David Giles for his excellent supervision of my work in researching for and preparing this thesis. His suggestions, encouragement, constructive criticism and timely advice have been invaluable. I am especially grateful to him for initially suggesting the topic of research and guidance throughout this course of study.

Financial support during this period of study was provided through the generosity of the New Zealand Vice-Chancellor's Committee (formerly the University Grants Committee), the William and Ina Cartwright Scholarship Fund and the William Georgetti Scholarship Fund, to whom I am grateful. In this regard I also wish to thank Alan Woodfield (Head of Economics Department) for providing me with further assistance through part-time teaching assignments and to the Department of Economics for providing me with the means to go to the 1991 Australasian Meeting of the Econometric Society, from which I benefitted enormously.

I am also grateful to Dr Judith Giles for her helpful comments on earlier drafts of parts of this thesis and for her support and encouragement, along with that of Alan Wan, Robin Harrison, Alan Woodfield, Paul Walker and Ewen McCann. To Karilyn Smith I owe a debt of gratitude for her speedy and accurate typing and her ability to decipher my writing.

Finally I wish to thank my family and friends, particularly my parents, for their loving and prayerful support and, most of all, my wife, Karina, for her prayerful encouragement, support, patience, love and for believing in me. I wish to dedicate this thesis to her and to our son Steven.

Contents

	Page
Abstract	ix
Chapter One Introduction	
1.1 Introductory Comments	1
1.2 Some Definitions and Background Analysis	3
1.3 An Outline of the Thesis	8
Chapter Two Pre-testing in the General Linear Model	
2.1 Introduction	11
2.2 Component Estimators	13
2.2.1 Necessary and Sufficient Conditions for the Restricted Estimators to Dominate the Unrestricted Estimators	16
2.3 The Pre-Test and Pre-Test Estimators of the Coefficient Vector and Prediction Vector	17
2.3.1 Optimal Critical Value of the Pre-Test When Estimating the Coefficient or Predictor Vectors	23
2.4 Pre-Test Estimators of the Scale Parameter	27
2.4.1 Optimal Critical Values of the Pre-Test When Estimating the Scale Parameter	30
2.5 Pre-Test Estimation in a Mis-Specified Model	31
2.6 Concluding Remarks	41
Chapter Three The Effect of a Mis-Specified Error Covariance Matrix on a Pre-Test for Linear Restrictions	
3.1 Introduction	44
3.2 Heteroscedastic Errors in the Linear Regression Model	45
3.2.1 The Effects of Uncorrected Heteroscedasticity on the Chow Test	48

	Page
3.3	Autocorrelated Errors in the Linear Regression Model 52
3.3.1	The Effects of Uncorrected Autocorrelation on a Pre-Test for Linear Restrictions 55
3.4	Concluding Remarks 58
Chapter Four	Pre-Test Estimation of the Coefficient Vector in a Linear Model with a Mis-Specified Error Covariance Matrix
4.1	Introduction 59
4.2	Properties of the Component Estimators and the Test Statistic 60
4.3	The Bias and Risk of the Pre-Test Estimator 64
4.4	Numerical Evaluation 66
4.4.1	The Models 66
4.4.2	Numerical Results 69
4.5	Conclusion 75
Appendix 4A	Proofs 78
Appendix 4B	The Regressor Data 88
Chapter Five	Pre-Test Estimation of the Coefficient and Predictor Vectors in a Model with a Mis-Specified Error Covariance Matrix and Excluded Relevant Regressors
5.1	Introduction 89
5.2	Properties of the Component Estimators and the Test Statistic 89
5.3	The Bias and Risk of the Pre-Test Predictor and the Pre-Test Estimator 93
5.4	Numerical Evaluation 96
5.4.1	Numerical Considerations 96
5.4.2	Numerical Results 98
5.5	Conclusion 107
Appendix 5A	Proofs 109

	Page
Chapter Six Pre-Test Estimation of the Scale Parameter in a Linear Model With a Mis-Specified Error Covariance Matrix	
6.1 Introduction	118
6.2 Properties of the Component Estimators	118
6.3 The Bias and Risk of the Pre-Test Estimators	122
6.4 Numerical Evaluation	126
6.4.1 The Model	126
6.4.2 Numerical Results	129
6.5 Conclusion	137
Appendix 6A Proofs	138
Chapter Seven Pre-Test Estimation of the Scale Parameter in a Linear Model With a Mis-Specified Error Covariance Matrix and Excluded Relevant Regressors	
7.1 Introduction	157
7.2 Properties of the Component Estimators	158
7.3 The Bias and Risk of the Pre-Test Estimators	160
7.4 Numerical Evaluation	163
7.4.1 The Model	163
7.4.2 Numerical Results	164
7.5 Conclusion	175
Appendix 7A Proofs	177
Chapter Eight The Robustness of Optimal Pre-Test Critical Values in a Linear Model with a Mis-Specified Error Covariance Matrix	
8.1 Introduction	181
8.2 Optimality Criteria	182
8.3 Numerical Results	187
8.4 Conclusion	194
Chapter Nine Concluding Remarks	195
References	202

List of Figures

	Page
Figure 2.1: Predictor Risk Functions, Well Specified Model	21
Figure 2.2: Scale Parameter Estimator Risk Functions, ML Estimators, Well Specified Model	28
Figure 2.3: Scale Parameter Estimator Risk Functions, LS Estimators, Well Specified Model	28
Figure 2.4: Scale Parameter Estimator Risk Functions, MS Estimators, Well Specified Model	29
Figure 4.1: Coefficient Estimator Risk Functions, Well Specified Model	71
Figure 4.2: Coefficient Estimator Risk Functions, MA(1) errors; $\rho = -0.9$	71
Figure 4.3: Coefficient Estimator Risk Functions, Well Specified Model	73
Figure 4.4: Coefficient Estimator Risk Functions, AR(1) Errors, $\rho = -0.9$	73
Figure 4.5: Coefficient Estimator Risk Functions, Well Specified Model	74
Figure 4.6: Coefficient Estimator Risk Functions, Heteroscedastic Errors (Model B): $\Psi = 10.0$	74
Figure 5.1: Pre-Test Power Functions, Excluded Regressor, Scalar Error Covariance Matrix	99
Figure 5.2: Predictor Risk Functions, Well Specified Model	100
Figure 5.3: Predictor Risk Functions, Excluded Regressor, Scalar Error Covariance Matrix	101
Figure 5.4: Coefficient Estimator Risk Functions, Well Specified Model	101
Figure 5.5: Coefficient Estimator Risk Functions, Excluded Regressor, Scalar Error Covariance Matrix	102
Figure 5.6: Coefficient Estimator Risk Functions, Excluded Regressor, Scalar Error Covariance Matrix	102
Figure 5.7: Predictor Risk Functions, Well Specified Model	104

	Page
Figure 5.8 Predictor Risk Functions, Excluded Regressor Scalar Error Covariance Matrix	104
Figure 5.9: Coefficient Estimator Risk Functions, No Excluded Regressors, AR(1) Errors, $\rho = 0.9$	106
Figure 5.10: Coefficient Estimator Risk Functions, Excluded Regressor, AR(1) Errors, $\rho = 0.9$	106
Figure 6.1: Scale Parameter Estimator Risk Functions, LS Estimators, Well Specified Model	130
Figure 6.2: Scale Parameter Estimator Risk Functions, LS Estimators, $\Psi = 10.0$	131
Figure 6.3: Scale Parameter Estimator Risk Functions, LS Estimators, $\Psi = 0.1$	132
Figure 6.4: Scale Parameter Estimator Risk Functions, ML Estimators, Well Specified Model	133
Figure 6.5: Scale Parameter Estimator Risk Functions; ML Estimators, $\Psi = 10.0$	134
Figure 6.6: Scale Parameter Estimator Risk Functions, ML Estimators, $\Psi = 0.1$	134
Figure 6.7: Scale Parameter Estimator Risk Functions, ML Estimators, Well Specified Model	135
Figure 6.8: Scale Parameter Estimator Risk Functions, MS Estimators, $\Psi = 10.0$	136
Figure 6.9: Scale Parameter Estimator Risk Functions, MS Estimators, $\Psi = 0.1$	136
Figure 7.1: Scale Parameter Estimator Risk Functions, LS Estimators, Well Specified Model	165
Figure 7.2: Scale Parameter Estimator Risk Functions, ML Estimators, Well Specified Model	166
Figure 7.3: Scale Parameter Estimator Risk Functions, MS Estimators, Well Specified Model	166
Figure 7.4: Scale Parameter Estimator Risk Functions, LS Estimators, Excluded Regressors, $\Psi = 1.0$	167
Figure 7.5: Scale Parameter Estimator Risk Functions, ML Estimators, Excluded Regressor, $\Psi = 1.0$	167

	Page
Figure 7.6: Scale Parameter Estimator Risk Functions, MS Estimators, Excluded Regressor, $\Psi = 1.0$	168
Figure 7.7: Scale Parameter Estimator Risk Functions, LS Estimators, Excluded Regressor, $\Psi = 1.0$	169
Figure 7.8: Scale Parameter Estimator Risk Functions, LS Estimators, Excluded Regressor, $\Psi = 0.01$	171
Figure 7.9: Scale Parameter Estimator Risk Functions, ML Estimators, Excluded Regressor, $\Psi = 0.01$	172
Figure 7.10: Scale Parameter Estimator Risk Functions, MS Estimators, Excluded Regressor, $\Psi = 0.1$	173
Figure 7.11: Scale Parameter Estimator Risk Functions, MS Estimators, Excluded Regressor, $\Psi = 0.1$	174
Figure 7.12: Scale Parameter Estimator Risk Functions, MS Estimators, Excluded Regressor, $\Psi = 10.0$	174
Figure 8.1: Coefficient Estimator Risk Functions, Well Specified Model	184
Figure 8.2: Coefficient Estimator Risk Functions, Well Specified Model	192

List of Tables

	Page
Table 8.1: Optimal Pre-Test Critical Values and Sizes, $T = 18, K = 2, J = 1$	189
Table 8.2: Optimal Pre-Test Critical Values and Sizes, $T = 18, K = 3, J = 1$	190
Table 8.3: Optimal Pre-Test Critical Values and Sizes, $T = 21, K = 5, J = 3$	191
Table 8.4: Optimal Pre-Test Critical Values and Sizes, $T = 19, K = 3, J = 2$	191
Table 8.5: The Effects of a Change in Sample Size	193

ABSTRACT

This thesis considers some finite sample properties of a number of preliminary test (pre-test) estimators of the unknown parameters of a linear regression model that may have been mis-specified as a result of incorrectly assuming that the disturbance term has a scalar covariance matrix, and/or as a result of the exclusion of relevant regressors. The pre-test itself is a test for exact linear restrictions and is conducted using the usual Wald statistic, which provides a Uniformly Most Powerful Invariant test of the restrictions in a well specified model. The parameters to be estimated are the coefficient vector, the prediction vector (*i.e.* the expectation of the dependent variable conditional on the regressors), and the regression scale parameter. Note that while the problem of estimating the prediction vector is merely a special case of estimating the coefficient vector when the model is well specified, this is not the case when the model is mis-specified.

The properties of each of these estimators in a well specified regression model have been examined in the literature, as have the effects of a number of different model mis-specifications, and we survey these results in Chapter Two. We will extend the existing literature by generalising the error covariance matrix in conjunction with allowing for possibly excluded regressors. To motivate the consideration of a nonscalar error covariance matrix in the context of a pre-test situation we briefly examine the literature on autoregressive and heteroscedastic error processes in Chapter Three.

In Chapters Four, Five, Six, and Seven we derive the cumulative distribution function of the test statistic, and exact formulae for the bias and risk (under quadratic loss) of the unrestricted, restricted and pre-test estimators, in a model with a general error covariance matrix and possibly

excluded relevant regressors. These formulae are data dependent and, to illustrate the results, are evaluated for a number of regression models and forms of error covariance matrix. In particular we determine the effects of autoregressive errors and heteroscedastic errors on each of the regression models under consideration.

Our evaluations confirm the known result that the presence of a non scalar error covariance matrix introduces a distortion into the pre-test power function and we show the effects of this on the pre-test estimators. In addition to this we show that one effect of the mis-specification may be that the pre-test and restricted estimators may be strictly dominated by the corresponding unrestricted estimator even if there are no relevant regressors excluded from the model. If there are relevant regressors excluded from the model it appears that the additional mis-specification of the error covariance matrix has little qualitative impact unless the coefficients on the excluded regressors are small in magnitude or the excluded regressors are not correlated with the included regressors.

As one of the effects of the mis-specification is to introduce a distortion into the pre-test power function, in Chapter Eight we consider the problem of determining the optimal critical value (under the criterion of minimax regret) for the pre-test when estimating the regression coefficient vector. We show that the mis-specification of the error covariance matrix may have a substantial impact on the optimal critical value chosen for the pre-test under this criterion, although, generally, the actual size of the pre-test is relatively unaffected by increasing degrees of mis-specification.

Chapter Nine concludes this thesis and provides a summary of the results obtained in the earlier chapters. In addition, we outline some possible future research topics in this general area.

CHAPTER ONE

INTRODUCTION

1.1 Introductory Comments

It is part of the nature of economics that there is always some element of doubt regarding the exact specification of a particular model, the accuracy of the available data, and the true data generating process involved. In economics it is not generally possible to perform or repeat an experiment in order to gain further information about the processes involved or to learn of the distribution of stochastic elements of the data. Economic models are, therefore, built on the basis of economic theory and assumptions which may, or may not, be borne out by the particular data available. Frequently there are conflicting model specifications proposed by economic theory and one of the purposes of econometric estimation of the model may be to provide some evidence as to which, if any, of the proposed specifications or restrictions is likely to be correct.

The imposition of restrictions on a statistical model may lead to an increase in the precision of the estimators applied to the model, i.e. a reduction in the variance, or covariance of the estimators, but it may also lead to a situation in which the estimators' biases increase or they may exhibit a lack of desirable statistical properties if the restrictions are not valid. A common way of determining which of several, theoretically valid, specifications is the most appropriate is through the application of a statistical tests. On the basis of assumptions regarding the distribution of the stochastic element of an economic process we may test for the significance of one or more regressors, for autocorrelation or

heteroscedasticity in the error term, non-normality of the regression residuals or the order of integration of the data, among other things.

On the basis of a test, or a number of tests, the applied worker goes from the initial model specification to a 'final' specification to which the appropriate estimators are applied. This procedure is termed pre-testing and it is clear that, as the estimators applied to the final model are, in part, determined by the outcome of prior statistical tests, the results of estimation are, in fact, conditional on the preliminary tests that have been undertaken. Such estimators are termed pre-test estimators and, in general, their properties differ from the properties of the estimators that would have been applied to the model had pre-testing not taken place.

There is a large body of literature dealing with the properties of pre-test estimators, particularly in the context of pre-testing linear restrictions in the standard linear regression model with independent, identically distributed $N(0, \sigma^2)$ errors. Typically, however, it is assumed that the only uncertainty present in the model under consideration relates to the validity of the restrictions being tested. This assumption may be violated for a number of reasons such as; the unavailability of data in an appropriate form, incorrect theory may lead to an initial model that is over or under specified in terms of which regressors are included, the regression disturbances may not be normally distributed or may not be independently or identically distributed.

A number of studies have examined the effects of various model mis-specifications such as the exclusion of relevant regressors or non-normal disturbances on the properties of pre-test estimators and have shown that, in a mis-specified model, the estimators' properties may be markedly different from their properties in a well-specified model. As econometricians work

with models whose correct specification is unknown, it is necessary to investigate the effects on the estimators used when there are departures from the usual assumptions made prior to estimation. In this thesis we contribute to this literature by considering the consequences of two particular mis-specifications which may occur in pre-testing exact linear restrictions in the classical linear regression model. These mis-specifications, a departure from the assumption of independently, identically distributed errors and a mis-specification of the regressor matrix (in that there are relevant regressors excluded from the set of regressors), may or may not occur simultaneously. We investigate the finite sample properties of the pre-test itself and the estimators under consideration, in a situation in which the fitted model is the classical linear model.

1.2 Some Definitions and Background Analysis

The fitted model under consideration in this thesis is the classical linear model described by

$$y = X\beta + \epsilon, \tag{1.2.1}$$

where y is a $T \times 1$ random vector of observations on the dependent variable, X is a $T \times K$ non-stochastic matrix of full rank which contains T observations on K explanatory, regressor, variables, $K < T$, β is a $K \times 1$ vector of unobservable coefficients and ϵ is a $T \times 1$ random disturbance vector. In the fitted model ϵ is assumed to be $N(0, \sigma^2 I_T)$ distributed, where σ^2 is an unknown scale parameter.

It is the assumption of $N(0, \sigma^2 I_T)$ errors that we will relax by considering an error term distributed as a $N(\xi, \sigma^2 \Omega)$ random variable where ξ is a $T \times 1$ non-stochastic vector and Ω is a general positive definite symmetric matrix. The ξ vector may represent the effect of relevant regressors

incorrectly excluded from the fitted model if we define $\xi = Z\gamma$, where Z is a $T \times K_z$ non-stochastic matrix of full rank which contains T observations on K_z excluded regressors, $K_z < T$, and γ is a $K_z \times 1$ vector of coefficients. Note that, if $\gamma = 0$ and $\Omega = I_T$ the fitted model is, in fact, well specified.

The coefficient vector, β , and the scale parameter, σ^2 , are unknown and it is the objective, in fitting the model, 1.2.1, to the data, to obtain estimates of them. The decision rules applied to the data to estimate β or σ^2 are termed estimators. As the data themselves are random, it follows that the estimators used will give rise to estimates of β and σ^2 which are realisations of a stochastic process and will not, in general, be equal to their actual, but unknown, values.

Let us denote the actual value of a parameter vector of interest as μ and let $D(y)$ denote a decision rule or estimator of μ based on the value of the random dependent variable y . The loss involved in using $D(y)$ rather than μ depends on the true value of μ , the particular decision rule applied, and the data, and is denoted $L(D(y), \mu)$. $L(D(y), \mu) \geq 0$ by definition with $L(D(y), \mu) = 0$ if and only if $D(y) = \mu$. As y is random the loss associated with $D(y)$ is also random and hence it is common to consider the risk, or expected loss, of an estimator as a measure of its goodness of fit. The risk function is denoted¹ $\rho(D(y), \mu) = E[L(D(y), \mu)]$. Risk under quadratic, or squared error loss, $\rho(D(y), \mu) = E[(D(y), \mu)'(D(y), \mu)]$, is frequently used in the literature as it has a number of desirable properties, such as its ease of use and simplicity, and the fact that it penalizes errors in estimation based on (squared) magnitude regardless of sign, and hence negative errors

¹ This expectation of the loss function is taken with respect to the distribution of y .

will not cancel out positive errors. Other symmetric loss functions include absolute error loss² and weighted squared error loss.³ In some instances, an investigator may wish to weight positive and negative errors differently and in such a case an asymmetric loss function, such as the Linex loss function proposed by Varian (1975) may be used.⁴

Risk under squared error loss can be written $\rho(D(y), \mu) = \text{tr} \left[\text{MSE}(D(y), \mu) \right]$ where $\text{MSE}(D(y), \mu)$ is the Matrix Mean Squared Error, or risk matrix of $D(y)$ given μ and $\text{tr}(\cdot)$ is the trace operator. $\text{MSE}(D(y), \mu)$ measures the cross product of differences between the estimators $D(y)$ of μ and the true value of μ and is defined $\text{MSE}(D(y), \mu) = E \left[(D(y) - \mu)(D(y) - \mu)' \right]$. It is also true that $\rho(D(y), \mu) = \text{tr} \left[\text{covar}(D(y)) + B(D(y), \mu)B(D(y), \mu)' \right]$, where $B(D(y), \mu)$ is the estimator's bias, and can therefore be interpreted as representing the bias/variance trade-off inherent in estimation. For these reasons it is risk under quadratic loss that we consider in this thesis.

The criterion of choosing a decision rule so as to minimize risk or expected loss is limited in application as frequently there will be no estimator which has minimum risk for each and every feasible value of μ . Hence other criteria may be considered in deciding which decision rule to apply. These include:

unbiasedness. An estimator $D(y)$ of μ is unbiased if, on average, it is equal to μ , i.e. $B(D(y), \mu) \equiv E(D(y), \mu) - \mu = 0$.

² Work currently underway by Giles and Lieberman (1992b) considers pre-testing in terms of risk under absolute error loss.

³ See, for example, Judge and Bock (1978, pp.30-33).

⁴ See Giles and Giles (1991) for an example of the use of a Linex loss function in a pre-test situation.

admissibility. An estimator $D(y)$ of μ is admissible if there is no other estimator, $\hat{D}(y)$, such that $\rho(\hat{D}(y), \mu) \leq \rho(D(y), \mu)$ for all possible values of μ with $\rho(\hat{D}(y), \mu) < \rho(D(y), \mu)$ for at least one possible value of μ . If an estimator $D(y)$ is inadmissible it is said to be strictly dominated.

linearity. An estimator $D(y)$ of a $K \times 1$ vector μ is linear if $D(y) = Ay$ for some non-stochastic $K \times T$ matrix A .

efficient or best unbiased. If an unbiased estimator $D(y)$ of μ has a covariance matrix no "greater" than the covariance matrix of any other unbiased estimator $\hat{D}(y)$ of μ in the sense that $\text{Covar}(\hat{D}(y)) - \text{Covar}(D(y))$ is a positive semi-definite matrix, it is said to be efficient or best unbiased.

maximum likelihood. An estimator $D(y)$ is said to be the maximum likelihood estimator, MLE, if $D(y)$ represents the most likely value of μ given the assumptions of the model and the data. That is to say that the estimator $D(y)$ maximises the likelihood function $\mathcal{L}(y, D(y))$, where $\mathcal{L}(y, \mu)$ is the joint density function of the dependent variable.

For further discussion and references relating to desirable properties of an estimator see Judge and Bock (1978), Judge *et al.* (1985), Harvey (1990) and Gujarati (1988), among others.

It is well known (Aitken (1934), David and Neyman (1938)) that, given the assumptions of the classical linear model described by 1.2.1 above, the Best Linear Unbiased Estimator, BLUE, of the coefficient vector, β , is the Ordinary Least Squares, OLS, estimator

$$b = (X'X)^{-1}X'y. \quad 1.2.2$$

The OLS estimator is also the minimum MSE and ML estimator, given the assumption of Normally distributed disturbances.

The minimum MSE and ML estimators of the scale parameter, σ^2 , are defined as $s_{MS}^2 = (T+2-K)^{-1}(y-Xb)'(y-Xb)$ and $s_{ML}^2 = T^{-1}(y-Xb)'(y-Xb)$ respectively. These estimators are asymptotically unbiased, that is, unbiased if T is infinitely large, but are biased in finite samples. An unbiased estimator of the scale parameter is the OLS estimator $s_{LS}^2 = (T-K)^{-1}(y-Xb)'(y-Xb)$.

In addition to the sample information contained in 1.2.1, there may be some non-sample prior information suggested by economic theory relating to the specification of the model. In the classical linear model this information often may be expressed in the form of J ($\leq K$) linear restrictions

$$R\beta = r, \quad 1.2.3$$

where R is a known $J \times K$ non-stochastic matrix of full rank and r is a known $J \times 1$ vector of constants. If the restrictions 1.2.3 are valid and are imposed on the regression model, the BLUE, minimum MSE and ML estimator of the coefficient vector is the Restricted Least Squares, RLS, estimator

$$b^* = b + (X'X)^{-1}R' (R(X'X)^{-1}R')^{-1}(r-Rb), \quad 1.2.4$$

and the least squares, ML, and minimum MSE estimators of the scale parameter are defined as $s_{LS}^{*2} = (T+J-K)^{-1}(y-Xb^*)'(y-Xb^*)$, $s_{ML}^{*2} = T^{-1}(y-Xb^*)'(y-Xb^*)$, and $s_{MS}^{*2} = (T+J+2-K)^{-1}(y-Xb^*)'(y-Xb^*)$, respectively.

In general these estimators will have "smaller" covariance matrices than their unrestricted counterparts.⁵ Note, however, that if the restrictions

⁵ Here we mean that $\text{Covar}(b^*)$ is less than $\text{Covar}(b)$ in the sense that $\text{Covar}(b) - \text{Covar}(b^*)$ is a positive semi-definite matrix.

are not valid, the restricted estimators will generally be more biased and hence may have higher risks than the corresponding unrestricted estimators. It is common, therefore, to use a statistical test to determine whether or not the restrictions 1.2.3 are valid prior to imposing them. Given the assumptions of the model a Uniformly Most Powerful Invariant test of the hypothesis $H_0: R\beta = r$ vs $H_A: R\beta \neq r$ can be constructed using the Wald statistic

$$u = \frac{(Rb-r)' (R(X'X)^{-1}R')^{-1} (Rb-r)(T-K)}{(y-Xb)'(y-Xb)J}.$$

It is this test statistic and the unrestricted and restricted OLS, ML and MSE estimators defined above that form the components of the pre-test estimators we will consider.

1.3 An Outline of the Thesis

Having established in the previous section the basic problem with which we will be dealing with, and the criterion under which we will compare the estimators (that is, risk under quadratic loss), we go on, in Chapter Two, to provide a background for the research presented in this thesis by reviewing the literature that deals with the particular pre-test situation we are concerned with. The pre-test of interest is, as noted above, a pre-test for exact linear restrictions in the classical linear model. We survey the studies which consider the estimation of the coefficient vector, the prediction vector (that is, the expectation of the dependent variable conditional on the regressors), and the regression scale parameter.

The existing literature considers such pre-test estimation in the context of a model with a scalar error covariance matrix and, in surveying this literature, we will assume that this is the correct specification of the model. As we will later consider the effects of a mis-specified error

covariance matrix on the properties of the pre-test estimators and the optimal pre-test size, we pay particular attention to that portion of the literature which relates to the choice of an optimal pre-test size in a well-specified model and the effects of various forms of mis-specification on the properties of the estimators.

The review in Chapter Two considers pre-testing problems in the context of a well-specified error covariance matrix with no autoregressive or heteroscedastic process in the error term. To motivate the consideration of a mis-specified non-scalar error covariance matrix in the context of pre-testing in the general linear, in Chapter Three we review some of the reasons why the usual assumption of a scalar error covariance matrix might not be valid and survey the literature dealing with the effect of a mis-specified error covariance matrix on a pre-test for linear restrictions.

The next four chapters investigate some of the finite sample properties of the pre-test estimators under consideration. In Chapter Four we derive the bias and risk functions of the pre-test estimator of the coefficient vector in a model with a mis-specified error covariance matrix. Chapter Five also considers the estimators of the coefficient vector as well as unrestricted, restricted and pre-test predictors in a model with a mis-specified error covariance matrix in conjunction with the problem of mis-specification of the regressor matrix caused by excluding relevant regressors from the model. Chapters Six and Seven consider the same situation as Four and Five respectively in the context of estimating the regression scale parameter. To illustrate the results given in these chapters, we numerically evaluate the risk formulae for a variety of different models and regressor matrices.

In Chapter Eight we consider the determination of the optimal critical value for the pre-test, using a criterion of minimax regret, when estimating the coefficient vector in a model with a mis-specified error covariance matrix. In each chapter we assume that the researcher is unaware of the model mis-specification. Chapter Nine concludes the thesis.

CHAPTER TWO

PRE-TESTING IN THE GENERAL LINEAR MODEL

2.1 Introduction

In this chapter we survey the literature that deals with the issue of pre-testing for exact linear constraints on the parameters in the classical linear regression model.

The properties of the component (restricted and unrestricted) estimators and the test statistic are discussed in Section 2.2 and the particular pre-test estimators considered, namely ones of the coefficient vector, the prediction vector and of the regression scale parameter, are discussed in Sections 2.3 and 2.4.

In addition to the risk functions of these estimators, the issue of an optimal pre-test size is discussed, and the various "rules of thumb" proposed by different authors in considering different estimators and different optimality criteria are also discussed. The properties of these estimators generally depend on such things as the particulars of the data set under consideration, the form of the restrictions being tested and, of course, the correct specification of the regression model. However, it is true that econometricians work with models that may not, in fact, be correctly specified and a number of papers have considered the effects that various mis-specifications may have on the classical pre-test estimators. Such mis-specifications may take the form of excluded relevant regressors, included irrelevant regressors, a combination of these two (as in the proxy variable problem) and a mis-specified error distribution possibly in addition to a mis-specification of the regressor matrix.

This thesis extends the results found to be in the existing literature by considering a regression model in which the error term is not well

behaved, in the sense that the error covariance matrix is non-scalar, and by simultaneously allowing for a regressor matrix which may have been mis-specified by the exclusion of relevant regressors. Note that the situation we consider in later chapters is one in which a mis-specification of the model has occurred. We examine the effects on estimator risk and optimal critical values of incorrectly assuming a scalar error covariance matrix in a regression model. This is a different situation from the one in which a non-scalar error covariance matrix is suspected and a pre-test is used to determine whether or not there is significant evidence of heteroscedasticity or autocorrelation. Pre-test estimators of the latter type have also been examined in the literature (see Fomby & Guilkey (1978), King and Giles (1984), or Giles and Beattie (1987) for examples of pre-testing for autocorrelation or Greenberg (1980), who follows the approach of Taylor (1977, 1978) and Mandy (1984), for examples of pre-testing for heteroscedasticity). The question of applying, for example, an autocorrelation pre-test estimator in a model in which the error term is both autocorrelated and heteroscedastic has not been addressed, although Small (1991) has recently provided some evidence regarding the effect of heteroscedasticity on the size and power of various tests for autocorrelation.

This thesis considers the estimators which deal with the classical linear model as they are commonly used in applied situations and which have received a considerable amount of attention in the pre-test literature. Accordingly, it is on the literature dealing with these particular estimators that this survey focusses. There are many other pre-test problems which have been considered in the literature and mention is made of some of these in Section 2.6. It is beyond the scope of this survey to deal with these fully as they do not directly pertain to the actual problem considered in this

paper.¹

2.2 Component Estimators

A pre-test situation arises in econometrics when the outcome of a statistical test determines which estimator is to be used for the parameter(s) of interest. In the general linear model

$$y = X\beta + \varepsilon, \quad 2.2.1$$

where y is a $T \times 1$ vector of observations on a dependent variable, X is a $T \times K$ non-stochastic matrix of rank K , containing T observations on K explanatory variables and ε is a $T \times 1$ random vector, $\varepsilon \sim N(0, \sigma^2 I_T)$, there may be non-sample information about the unknown parameters, β , which can be represented in the form of J linear restrictions given by $R\beta = r$, where R is a $J \times K$ non-stochastic matrix of rank J and r is a $J \times 1$ non-stochastic vector.

An estimator which takes the non-sample information into account, a restricted estimator, will be more efficient than one which does not if the information embodied in the imposed restrictions is valid. However if the information is not valid the restricted estimator is likely to be more biased and may have a higher risk than an unrestricted estimator, despite being more precise. There is, therefore, some doubt as to whether or not to impose the restrictions on the model's parameter.

If the researcher chooses to ignore the non-sample information and estimate the coefficient vector, β , using the Maximum Likelihood approach, or the Least Squares approach, which is equivalent given the assumptions of the model, the (unrestricted) estimator applied is the ordinary least squares,

¹ See Giles and Giles (1993) for a survey of recent developments in the area of pre-testing in general. Further references may be found in the annotated bibliographies of Bancroft and Han (1977) and Han *et al.* (1988).

OLS, estimator, $b \equiv S^{-1}X'y$, where $S \equiv X'X$. The Gauss-Markov Theorem states that this estimator is the Best Linear Unbiased estimator. Because the y vector is normally distributed, b will also be normally distributed with a mean of $E(b) = \beta$ and a covariance matrix of $\text{covar}(b) = \sigma^2 S^{-1}$. The risk, under squared error loss, of this estimator is $\rho(b, \beta) = E[(b - \beta)'(b - \beta)] = \sigma^2 \text{tr}(S^{-1})$, where $\text{tr}(\cdot)$ refers to the trace of the matrix. Similarly the researcher could estimate the regression scale parameter, σ^2 , using $s_1^2 = (y - Xb)'(y - Xb) / (T + \Delta)$, where s_1^2 describes the Maximum Likelihood, s_{ML}^2 , Least Squares, s_{LS}^2 , or Minimum Mean Squared error, s_{MS}^2 , estimator according to whether $\Delta = 0$, $-K$ or $(2-K)$ respectively. Defining $M \equiv (I_T - XS^{-1}X')$ we have $s_1^2 = \epsilon' M \epsilon / (T + \Delta)$ and hence $(T + \Delta)s_1^2 / \sigma^2$ is distributed as a central $\chi_{(\nu)}^2$ random variable, where $\nu = T - K$. The bias and risk of s_1^2 are $B(s_1^2, \sigma^2) = E(s_1^2) - \sigma^2 = -\sigma^2(K + \Delta) / (T + \Delta)$ and $\rho(s_1^2, \sigma^2) = \sigma^4(2\nu + (K + \Delta)^2) / (T + \Delta)^2$ respectively. Therefore we see that the Least Squares estimator s_{LS}^2 is unbiased while the s_{ML}^2 and s_{MS}^2 estimators are biased estimators of σ^2 .

If the researcher imposes the restrictions and estimates the coefficient vector using Maximum Likelihood or, equivalently in this case, Least Squares, the estimator used is the restricted least squares, RLS, estimator, $b^* \equiv b + S^{-1}R'(RS^{-1}R')^{-1}(r - Rb)$. This estimator is also normally distributed with a mean of $E(b^*) = \beta - S^{-1}R'(RS^{-1}R')^{-1}\delta$, where $\delta = R\beta - r$ (the hypothesis error), and a covariance matrix of $\text{covar}(b^*) = \sigma^2(S^{-1} - C)$ where $C = S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$. The restricted estimator is unbiased if and only if the restrictions are valid, since this implies $\delta = 0$. As C is at least positive semi-definite the restricted estimator has a covariance matrix that is less than or equal to the covariance matrix of the unrestricted estimator, in the matrix sense. The risk of the restricted estimator is $\rho(b^*, \beta) = \sigma^2 \text{tr}(S^{-1} - C) + \delta' \eta' \eta \delta$, where $\eta = S^{-1}R'(RS^{-1}R')^{-1}$. As $\text{tr}(S^{-1} - C)$ is less than or equal to $\text{tr}(S^{-1})$ it is apparent that the restricted estimator has a lower risk than the unrestricted

estimator if the restrictions are valid. However as the hypothesis error increases the risk of the restricted estimator rises until the risk of the two estimators are the same and, as δ increases still further, the risk of the restricted estimator increases without bound. The question of how large the hypothesis error can be while still having the restricted estimator dominate the unrestricted estimator is addressed below.

Alternatively we may consider the problem of estimating the the mean forecast of y conditional on X , that is, of estimating $E(y)$, rather than estimating β . This is equivalent to estimating β in a well specified model in which the regressor matrix is orthonormal. The risk functions of the unrestricted and restricted estimators are $\rho(Xb, E(y)) = \sigma^2 K$ and $\rho(Xb^*, E(y)) = \sigma^2(K - J + 2\lambda)$ respectively in this case, where $\lambda \equiv \delta'(RS^{-1}R')^{-1}\delta/2\sigma^2$. If the restrictions are valid $\lambda = 0$ and hence the restricted estimator has a lower risk than the unrestricted estimator. However, the greater is the absolute error in the prior information, the greater is the value of λ . We can therefore view λ as a scalar measure of the validity of the restrictions.

The estimator of the scale parameter which results from imposing the restrictions is $s_1^{*2} = (y - Xb^*)'(y - Xb^*)/(T + \Gamma)$, where s_1^{*2} denotes the restricted Maximum Likelihood, s_{ML}^{*2} , the restricted Least Squares, s_{LS}^{*2} or the restricted Minimum Mean Squared Error, s_{MS}^{*2} , estimator as $\Gamma = 0$, $(J - K)$ or $(J + 2 - K)$ respectively. Note that the s_{MS}^{*2} estimator will not in fact be the estimator with the minimum mean squared error unless the restrictions are valid. As $s_1^{*2} = \epsilon^* M \epsilon^* / (T + \Gamma)$ it is apparent that $(T + \Gamma)s_1^{*2}/\sigma^2$ is distributed as a non-central $\chi^2_{(\nu + J; \lambda)}$ random variable. If the restrictions are true $\delta = 0$, $\lambda = 0$ and $(T + \Gamma)s_1^{*2}/\sigma^2$ is distributed as a central $\chi^2_{(\nu)}$ random variable. The bias and risk of the s_1^{*2} estimators are given by $B(s_1^{*2}, \sigma^2) = E(s_1^{*2}) - \sigma^2 = \sigma^2(J - K - \Gamma + 2\lambda)/(T + \Gamma)$ and $\rho(s_1^{*2}, \sigma^2) = \sigma^4(2(J + \nu + 4\lambda) + (J - \nu + 2\lambda)^2)/(T + \Gamma)^2$ respectively.

2.2.1 Necessary and Sufficient Conditions for the Restricted Estimators to Dominate the Unrestricted Estimators

As noted above, the restricted estimator of the β vector will be more precise than the unrestricted estimator. However, as the level of hypothesis error, δ , increases, the restricted estimator's bias increases without bound, as does the estimator's risk. For small absolute values of δ the restricted estimator may have a lower risk than the unrestricted estimator, despite being biased, as the effects of the bias on the risk may be offset by the effect of the greater estimator precision on the risk. Because the risks of both the OLS and the RLS estimators of β are data dependent the range of hypothesis error over which the RLS estimator dominates the OLS estimator is also data dependent. Wallace (1972) shows that the RLS estimator has lower risk than the OLS estimator if $\lambda \leq \frac{1}{2} d_L^{-1} \text{tr}(S^{-1}R'(RS^{-1}R')^{-1}RS^{-1})$, where d_L^{-1} is the inverse of the largest characteristic root of S^{-1} . Toro-Vizcarrondo and Wallace (1968) consider the related problem of determining the range of λ over which the MSE matrix of the RLS estimator is less (in the matrix sense) than the MSE matrix of the OLS estimator. That is, they require $\text{MSE}(b, \beta) - \text{MSE}(b^*, \beta) = E[(b - \beta)(b - \beta)'] - E[(b^* - \beta)(b^* - \beta)'] = \sigma^2 C - \eta \delta \delta' \eta'$ to be positive semi definite. They show that a necessary and sufficient condition for this to be the case is that $\lambda \leq \frac{1}{2}$. This condition has obvious appeal as it is independent of the data.

In the case of estimating the mean forecast of y , conditional on X , it is trivial to show that the risk of the restricted predictor is less than the risk of the unrestricted predictor iff $\lambda \leq \frac{J}{2}$. See Judge and Bock (1978) for a discussion and further details.

Similarly there is a range of λ over which the restricted estimators of σ^2 dominate their unrestricted counterparts. The boundaries of this range vary depending on which of the ML, LS or MS estimators is being considered.

It is straightforward to show that

$$\rho(s_{ML}^{*2}, \sigma^2) \leq \rho(s_{ML}^2, \sigma^2) \text{ iff } \lambda \leq -1 + (4\nu^2 + 2J\nu(\nu+J))^{1/2}/2\nu$$

$$\rho(s_{LS}^{*2}, \sigma^2) \leq \rho(s_{LS}^2, \sigma^2) \text{ iff } \lambda \leq J(\nu+J+2)/2(\nu+2)$$

$$\rho(s_{MS}^{*2}, \sigma^2) \leq \rho(s_{MS}^2, \sigma^2) \text{ iff } \lambda \leq \left[K-J-2+((J-K+2)^2 - J(J-2K+2))^{1/2} \right]/2.$$

Therefore, as is the case with estimating the coefficient vector, the restricted estimators may have a lower risk than the unrestricted estimators even if the restrictions are not valid. This is because the increased efficiency of the restricted estimators, compared to the unrestricted estimators, offsets to some extent the effect on estimator risk of the increase in bias introduced by imposing restrictions that are not exactly correct. However as δ and λ increase in magnitude the restricted estimator's biases and risks increase without bound. Given that the true value of λ is unknown the applied researcher will have some doubt as to whether or not the prior information should be used or ignored. It is in response to this uncertainty that a statistical test may be used to determine the validity of the proposed restrictions.

2.3 The Pre-test and Pre-Test Estimators of the Coefficient Vector and Prediction Vector

The prior information may be written as a hypothesis to be tested, $H_0: R\beta = r$ ($\delta=0$) vs $H_A: R\beta \neq r$ ($\delta \neq 0$) and the validity of this hypothesis can be tested using the Wald statistic

$$u = \frac{(Rb-r)' (RS^{-1}R')^{-1} (Rb-r)(T-K)}{(y-Xb)'(y-Xb)} \quad 2.3.1$$

Given the assumptions of the model, this statistic has a non-central $F'_{(J,\nu;\lambda)}$ distribution, with non-centrality parameter, λ , as defined above. Under the

null hypothesis $\delta = 0$, therefore $\lambda = 0$ and the statistic has a central $F_{(J,\nu)}$ distribution. For a given test size, α , the critical value for the test, c , is such that $\Pr.(u \leq c | \lambda=0) = (1-\alpha)$ and the hypothesis is rejected if the value of the test statistic is greater than this critical value. It is well known (e.g., Toro-Vizcarrondo and Wallace (1968)) that this is a Uniformly Most Powerful Invariant, UMPI, size- α test of the restrictions.

If the hypothesis is rejected, i.e., if $u > c$, the unrestricted estimator is subsequently used to estimate the parameters of the model, otherwise the restricted estimator is used. Therefore the estimator actually applied to the data is neither the naive (i.e., unpre-tested) restricted, nor the naive unrestricted estimator but is in fact a Pre-Test Estimator, PTE, which, in the case of estimating the coefficient vector, can be written

$$\hat{\beta} = \begin{cases} b; & \text{if } u \geq c, \\ b^*; & \text{if } u < c \end{cases}$$

or

$$\hat{\beta} = b^* I_{(0, c]}(u) + b I_{(c, \infty)}(u),$$

where

$$I_{(y, z]}(x) = \begin{cases} 1; & \text{if } x \in (y, z] \\ 0; & \text{else} \end{cases}$$

The properties of this PTE will vary with the data, the form of the restrictions and the size of the pre-test. In particular the properties of the PTE will differ from the properties of the two component estimators, b and b^* , though frequently this difference is not taken into account in an applied situation.

Pre-test estimators of this form were first considered by Bancroft (1944) whose simple regression model involved only two regressor variables and a single zero restriction on one coefficient. That is,

$$y = x_1\beta_1 + x_2\beta_2 + \varepsilon \quad 2.3.2$$

where y , x_1 , and x_2 are $T \times 1$ vectors of observations measured as deviations from their respective sample means. x_1 and x_2 are non-stochastic and are scaled such that they have unit variances and correlation coefficient ρ . The restriction to be tested is described by $H_0: \beta_2 = 0$ vs $H_A: \beta_2 \neq 0$ and Bancroft considers only the estimators of β_1 . Thus the unrestricted estimator is denoted $b_{1.2}$ and is obtained through OLS estimation of 2.3.2 while the restricted estimator is denoted b_1 and is obtained through OLS estimation of

$$y = x_1\beta_1 + \varepsilon^*.$$

The pre-test estimator is described by $\hat{\beta}_1 = b_{1.2}I_{(c, \infty)}(u) + b_1I_{(0, c)}(u)$. Bancroft shows the bias of the PTE to be

$$\text{Bias}(\hat{\beta}_1, \beta) = \rho\beta_2 \left(1 - \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} Ix_0\left(\frac{n-3}{2}, \frac{3}{2}+i\right) \right)$$

where $a = (1-\rho^2)(I-1)\beta_2^2/1$, $x_0 = \frac{1}{\frac{c}{n-3}+1}$ where c is the pre-test critical value

and $Ix_0(.,.)$ is the incomplete beta function. He also notes that the PTE will be biased unless ρ and/or β_2 are equal to zero; that is, either the restriction is, in fact, valid, or the regressors are orthonormal. The Mean Squared error of this PTE is derived by Toro-Vizcarrondo (1968). Brook (1972, 1976) extends the problem by considering a regression model with multiple restrictions. He derives the MSE of the PTEs of both β and $E(y)$. The risk of the PTE of β with orthonormal regressors is also derived by Sclove *et al.* (1972) and Bock *et al.* (1973) extend this work by deriving the risk of the PTE in the general case. In a survey article Wallace (1977) summarises the literature dealing with the Bancroft pre-test model and the PTE of $E(y)$. See Judge and Bock (1978), Judge *et al.* (1985) and Giles and Giles (1993) for further references and discussion.

The risk of the PTE of $E(y)$, i.e., the pre-test predictor, or, equivalently, the risk of the PTE of β when X is orthonormal, is

$$\rho(X\hat{\beta}, E(y)) = \sigma^2 (K + 4\lambda - Kh(2,0) - 2\lambda h(4,0)),$$

and the risk of the PTE of β in the general case is

$$\begin{aligned} \rho(\hat{\beta}, \beta) &= \sigma^2 \text{tr}(S^{-1}) - \sigma^2 \text{tr}(C)h(2,0) \\ &\quad - \delta' \eta' \eta \delta (h(4,0) - h(2,0)), \end{aligned}$$

$$\text{where } h(i,j) = \text{Pr.} \left[\frac{\chi^2_{(j+1;\lambda)}}{\chi^2_{(\nu+j)}} < \frac{cJ}{\nu} \right].$$

Giles and Srivastava (1993) take the analysis further by deriving the exact distribution of a simple pre-test estimator. The model they consider is that used by Bancroft (1944) described by 2.3.2 in which an exclusion restriction on one coefficient in a two regressor model is tested. The pre-test estimator is described in 2.3.3 above.

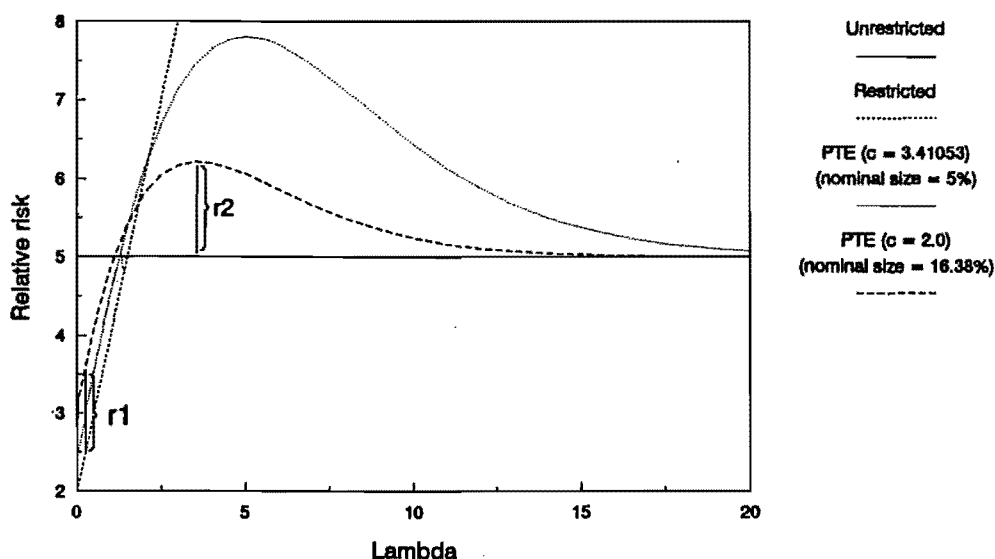
Based on their numerical evaluations, Giles and Srivastava note that the true confidence level associated with an interval estimate of β_1 differs from the nominal confidence level, in general. In the region of the null the confidence level of the PTE is less than the confidence level of the unrestricted estimator. However, as the restriction becomes increasingly false the confidence level of the PTE is greater than the true confidence level of the restricted estimator, which declines to zero. Giles and Srivastava note that the true confidence level of the pre-test estimator is always less than the nominal level, implying that confidence intervals constructed after applying restrictions to the model should be interpreted cautiously.

The risk functions of the OLS predictor, Xb , the RLS predictor, Xb^* , and the Pre-test, PT, predictor, $X\hat{\beta}$, are shown in² Figure 2.1. It is apparent that Xb is dominated by both Xb^* and $X\hat{\beta}$ if the restrictions are valid, i.e., if $\lambda = 0$.

FIGURE 2.1: Predictor Risk Functions

Well Specified Model

Five regressors



Three restrictions

Sample size = 18

Also, because the pre-test will reject valid restrictions with a probability of $\alpha\%$, the PT prediction risk is greater than the RLS predictor risk in this case. As the test size increases there is a greater chance of rejecting valid prior information and hence the risk difference between the PT and the OLS predictors will decrease while the risk difference between the PT and RLS predictors will increase.

²

Although the risk functions of the unrestricted, restricted and pre-test estimators of the coefficient vector are data dependent, qualitatively their properties are the same as the risk functions of the corresponding predictors in a well specified model.

Conversely, if the restrictions are incorrect and λ is very large, the OLS predictor has the least risk and the risk of the PT predictor lies between the OLS and the RLS risk. As λ increases the pre-test is more likely to reject the restrictions and hence the unrestricted predictor is more likely to be chosen. As $\lambda \rightarrow \infty$ the probability of rejecting the restrictions (the power of the pre-test) tends to unity and therefore the PT predictor risk approaches the OLS risk from above.

As noted above, the RLS predictor has lower risk than the OLS predictor for values of $\lambda < \frac{J}{2}$, they have the same risk at $\lambda = \frac{J}{2}$ and the RLS predictor has greater risk than the OLS predictor when $\lambda > \frac{J}{2}$. Although the value of λ at which the PT and the OLS predictors have the same risk varies, depending on the critical value of the pre-test, it is straightforward to show that it occurs somewhere in the range $\frac{J}{4} \leq \lambda \leq \frac{J}{2}$. (See Judge and Bock (1978, pp.72-75) for a discussion on the bounds on λ for the equality of the risk functions.) Therefore the minimum risk predictor for $\lambda \in [0, \frac{J}{2})$ is the RLS predictor or, equivalently, the PT predictor with a test size of $\alpha = 0\%$, while the minimum risk predictor for $\lambda \in (\frac{J}{2}, \infty)$ is the OLS predictor or, equivalently, the PT predictor with a test size of $\alpha = 100\%$. In any case other than the trivial ones of $\alpha = 0\%$ or 100% the PT predictor's risk function increases monotonically as λ increases above 0, reaches a single maximum, and then decreases monotonically to the OLS predictor risk as $\lambda \rightarrow \infty$. This maximum of the PT predictor risk occurs to the right of the point at which the PT predictor risk equals the RLS risk, as is also illustrated in Figure 2.1.

Of the estimators considered, no one estimator strictly dominates the other two or is itself strictly dominated by one of the others.³ The PTE is never the minimum risk estimator and, over some part of the λ space, it is the estimator with the highest risk. Therefore, in a world of perfect information, the researcher would never choose to pre-test. In the absence of perfect information a researcher may choose to pre-test rather than ignore, possibly valid, prior information or impose, possibly wildly inaccurate, prior information. The question then arises as to what is an appropriate test size. Traditionally, pre-tests have been carried out at significance levels of $\alpha = 1\%$ or $\alpha = 5\%$. However such arbitrary sizes fail to take into account any features of the particular model under consideration. The smallest possible risk that would result from estimation of the model is given by $\rho(Xb^*, E(y))$ if $\lambda \leq \frac{J}{2}$ and $\rho(Xb, E(y))$ otherwise. As noted above the PTE risk will tend towards the RLS risk as α decreases or, in other words, as the critical value increases. This will mean that the PTE risk is closer to the minimum risk if $\lambda \leq \frac{J}{2}$, but further from it otherwise. The situation then is one in which a trade off must be made.

2.3.1 Optimal Critical Value of the Pre-Test When Estimating the Coefficient or Predictor Vectors

To determine the optimal test size some definition of optimality is required. One possible approach is to choose a critical value according to a

³ In fact there are estimators that do strictly dominate the OLS estimator, such as the Bayesian pre-test estimator (see, for example, Leonard (1977)) or the Stein rule estimator (see Judge and Bock (1978) for a discussion of this estimator and further references). Similarly, Cohen (1965) proves that, under squared error loss, this PTE is inadmissible (under certain conditions) and Sclove *et al.* (1972) propose another pre-test estimator which dominates the one considered here. Despite its inadmissibility, the PTE under discussion is frequently encountered in practice, so there is a genuine interest in its sampling properties under a variety of circumstances.

minimax risk criterion. That is, we seek to minimise the maximum risk of the PTE across all values of λ . The solution is, however, trivial in this case as the maximum risk of the PTE will always be at least as great as the risk of the OLS estimator and will only equal the OLS risk if $c = 0$ (or $\alpha = 100\%$). Therefore the optimal critical value under this criterion is $c = 0$ implying that the OLS estimator should be applied whenever there is any doubt at all about the validity of the prior information. Conversely if a minimum risk criterion is used the optimal critical value will be infinitely large implying that the RLS estimator should always be used even if the researcher has grave doubts about the validity of the prior information. See Wallace and Ashar (1972) and Bock *et al.* (1973) for further discussion.

An alternative suggestion, made by Toyoda and Wallace (1976) is to search for a critical value such that the area between the PTE risk function and the minimum risk boundary is minimised. This is equivalent to minimising the average regret of choosing the PTE estimator as opposed to the minimum risk estimator over $\lambda \in [0, \infty)$. Toyoda and Wallace use an iterative search procedure to compute the value of c that minimises $\int_0^\infty \left(\rho(\hat{X}\hat{\beta}, E(y)) - \min(\rho(Xb^*, E(y)), \rho(Xb, E(y))) \right) d\lambda$. They show that there is no one critical value that is optimal in this sense across all models regardless of degrees of freedom of the pre-test. For $J < 5$, that is less than five restrictions, they conclude that the optimal critical value is $c = 0$; otherwise the optimal critical value is roughly two.

Another possible criterion for determining the optimal critical value for the pre-test is that of minimax regret. Regret, at given values of λ and c , is defined as the difference between the risk of the PTE associated with that critical value and the minimum possible PTE risk across all critical values. The optimal critical value is the one which minimises the maximum

values of that regret. In the model under consideration this amounts to choosing the value of c which minimises $\max_{\lambda} \left(\rho(\hat{X}\beta, E(y)) - \min(\rho(Xb^*, E(y)), \rho(Xb, E(y))) \right)$. This is the approach used by Sawa and Hiromatsu (1973) who tabulate optimal critical values for a model with a single exclusion restriction. Brook (1976) generalises this work to a model with general restrictions. Brook considers the maximum regret associated with pre-testing given that $\lambda \leq \frac{J}{2}$, given by r_1 in Figure 2.1 and the maximum regret associated with pre-testing given that $\lambda \leq \frac{J}{2}$, given by r_2 in Figure 2.1. Clearly the maximum regret over $\lambda \in [0, \infty)$ must be either r_1 or r_2 . As the critical value increases, r_1 decreases while r_2 increases, and the converse occurs as the critical value decreases. Hence, the optimal critical value in this model is the one at which $r_1 = r_2$. In the case of estimating $X\beta$ (or the case of estimating β with orthonormal regressors) Brook finds that the optimal critical value is roughly two, regardless of the regression degrees of freedom. Clearly such a result has great practical appeal, however it does not carry over to the case of estimating β in a model where the regressors are not orthonormal, unless there is only one restriction being tested.

Although both Brook's (1976) and Toyoda and Wallace's (1976) optimal critical values are approximately invariant to changes in the degrees of freedom of the pre-test, the optimal size of the test which is associated with these critical values is not. The optimal size decreases as the denominator degrees of freedom increase, as we would expect given that the optimal critical value is roughly fixed, however in general it is apparent that, for small or moderate degrees of freedom, the optimal size is greater than the 5% level traditionally used for such tests.

In a later paper Brook and Fletcher (1981) address the problem of finding the optimal critical value when estimating β in a model with general

restrictions and general regressors using both Toyoda and Wallace's (1976) criterion and Brook's (1976) criterion. Brook and Fletcher scale X and reorder the model to form $y^* = X^*\beta^* + \varepsilon^*$, where $X^{*'}X^*$ is in autocorrelation form⁴ and the restrictions are $R^*\beta^* = (0, I_J)\beta^* = 0$. They show that, using the Toyoda and Wallace criterion, the optimal critical value is $c = 0$ when $t \equiv \text{tr}(R^*S^{*-1}R^*) \leq 4$, and otherwise the optimal critical value is approximately $c = \frac{(\nu)}{(\nu+2)} \left(1 + \frac{t}{J} \right)$. They also conjecture that, using the minimax regret criterion of Brook (1972, 1976), the optimal critical value is approximately⁵ $c = 1 + \frac{t}{J}$, for $J > 1$. Brook and Fletcher tabulate exact optimal critical values under both of these criteria and show that these approximations become more accurate the greater are⁶ J and $(T-K)$. These results provide an easily applicable rule of thumb to use in determining the optimal critical value under one or other of these criteria and of applying a PTE with desirable risk properties.

Brook and Fletcher point out that it could be argued that relatively more weight should be given to smaller values of λ when determining the optimal critical value as the presence of the prior information itself is some indication that λ is likely to be closer to 0 than might be supposed in the absence of the prior information. The criteria considered thus far have in effect treated λ as being equally likely to fall anywhere in the interval $[0, \infty)$. If more weight was to be given to the smaller values of λ then this

⁴ Brook and Fletcher do this by expressing the dependent and independent variables as deviations about their sample means and scaling so that $x_i'x_i = 1$, where x_i is the i 'th regressor.

⁵ Brook and Fletcher are unable to justify this but find the true optimal to be closely approximated by $c = 1 + \frac{t}{J}$ for the data sets they consider.

⁶ Note that if the regressors are orthonormal $t = J$ and therefore the optimal critical values correspond to those previously calculated by Toyoda and Wallace (1976) and Brook (1976).

would imply treating the optimal critical values they calculate as lower bounds on the appropriate critical value.⁷

2.4 Pre-Test Estimators of the Scale Parameter

We now turn our attention to the PTEs of the scale parameter,⁸ σ^2 . These estimators are constructed in the same way as the PTEs of the coefficient vector with

$$\hat{\sigma}_1^2 = \begin{cases} s_1^2; & \text{if } u \geq c \\ s_1^{*2}; & \text{if } u < c \end{cases}$$

or

$$\hat{\sigma}_1^2 = s_1^{*2} I_{(0, c]}(u) + s_1^2 I_{(c, \infty)}(u).$$

The risk functions of the estimators are derived and evaluated by Clarke *et al.* (1987a & b) and are

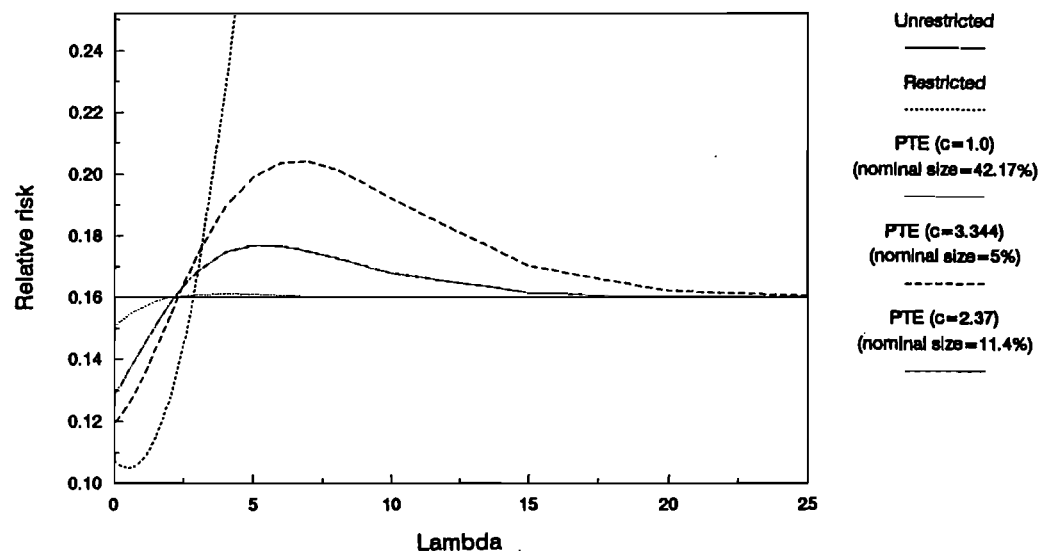
$$\begin{aligned} \rho(\hat{\sigma}_1^2, \sigma^2) &= \sigma^4 \left[1 + ((T+\Gamma)(T+\Delta))^{-2} \left(4\lambda(T+\Delta)^2 (\lambda h(8,0)) \right. \right. \\ &\quad + (J+2)h(6,0) + \nu h(4,0) - (T+\Gamma)h(4,0) \\ &\quad + \nu(\nu+2)(T+\Gamma)^2 - 2(T+\Gamma)(T+\Delta) \\ &\quad + (\nu(T+\Gamma) + \nu(\Delta-\Gamma)h(0,2) + J(T+\Delta)h(2,0)) \\ &\quad + J(T+\Delta)^2 (2\nu h(2,2) + (J+2)h(4,0)) \\ &\quad \left. \left. + \nu(\nu+2)(\Delta-\Gamma)(2T+\Delta+\Gamma)h(0,4) \right) \right]. \end{aligned}$$

⁷ This point had been previously made by Wallace (1977).

⁸ Note that this problem differs from that of estimating the standard error of estimate, σ , for which problem see Clarke (1986, 1990).

These functions depend on the data only through T , K , J and λ and are graphed in Figures 2.2, 2.3, and 2.4, which show examples of the ML, LS and MS estimators respectively.

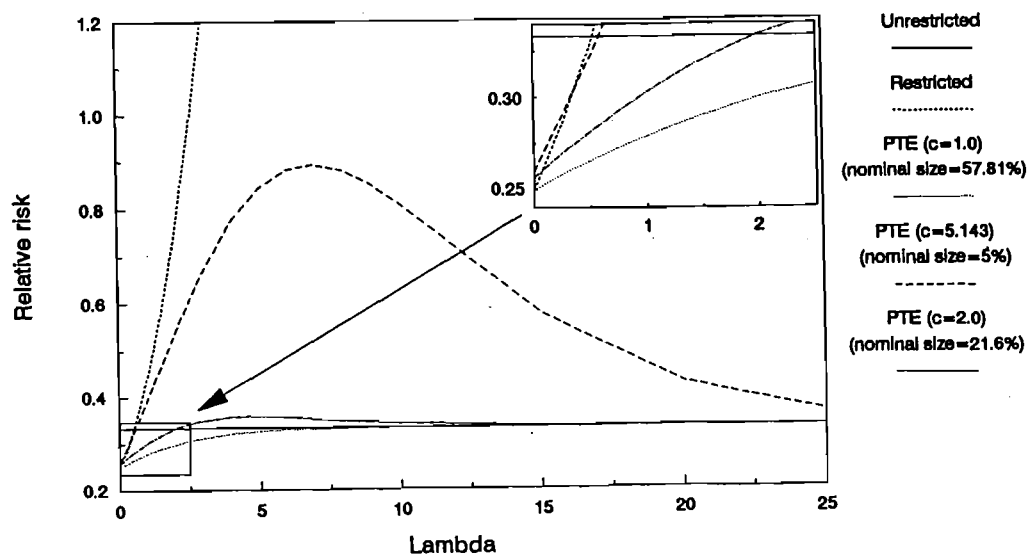
FIGURE 2.2: Scale Parameter Estimator Risk Fns.
ML Estimators, Well Specified Model
Three regressors



Application of Chow test

Sample size = 20

FIGURE 2.3: Scale Parameter Estimator Risk Fns.
LS Estimators, Well Specified Model
Two regressors



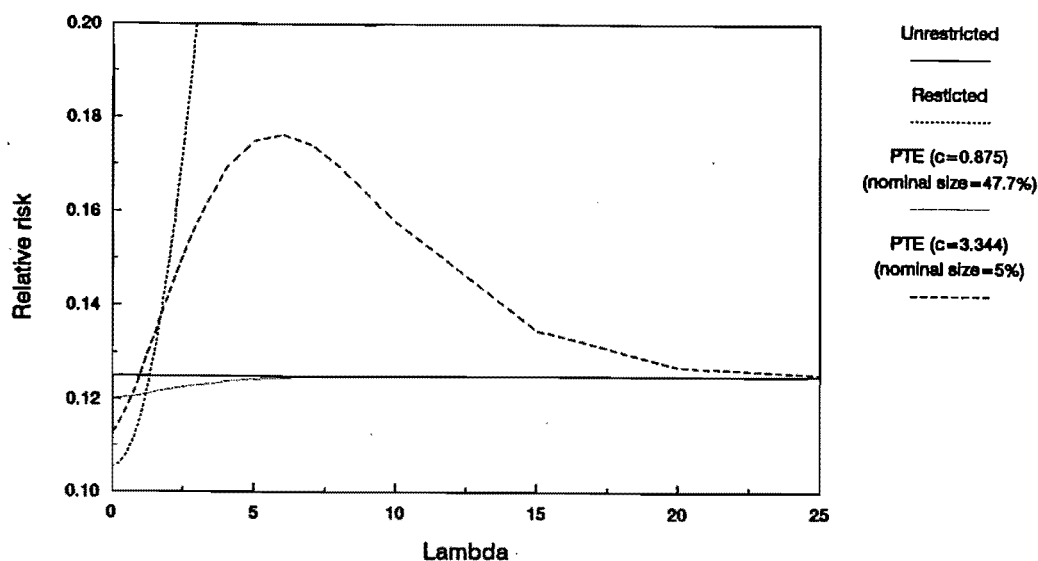
Application of Chow test

Sample size = 10

FIGURE 2.4: Scale Parameter Estimator Risk Fns.

MS Estimators, Well Specified Model

Three regressors



Application of Chow test

Sample size = 20

From these Figures it is apparent that, as in the case of the estimators of the coefficient vector, the PT risk approaches the risk of the restricted estimator as the critical value increases, while the converse occurs as the critical value decreases. In the case of the ML estimators the PTE is never the least risk estimator and is in fact the highest risk estimator over some part of the parameter space, this is analogous to the case of estimating the coefficient vector. However in the case of the LS estimators, Giles (1991a) shows that there exists a family of pre-test estimators which strictly dominate the unrestricted estimator. Her numerical results suggest that the PTE associated with a pre-test critical value of $c = 1$ strictly dominates both the restricted and unrestricted estimators when $J \leq 2$. In the case where $J > 2$ the minimum risk estimator is the restricted estimator for small values of λ and the PTE with $c = 1$ for larger values of λ .

Similarly, Ohtani (1988) considers the MS estimators and demonstrates numerically that there exists a family of PTEs which dominate the

unrestricted estimator. Ohtani proves that the minimum risk member of this family is the PTE which is associated with a choice of critical value of $c = \nu/(\nu+2)$ and that the PTE corresponding to the use of this critical value is the Stein (1964) estimator. This result is proved analytically by Gelfand and Dey (1988).

2.4.1 Optimal Critical Values of the Pre-Test When Estimating the Scale Parameter

With the exception of the LS estimators when $J \leq 2$, there is no one dominating estimator in these examples and therefore the issue of deciding an appropriate critical value for the pre-test is raised. In terms of minimising the PTE risk, Ohtani (1988) has shown that $c = \nu/(\nu+2)$ is the optimal for the $\hat{\sigma}_{MS}^2$ estimator and Giles (1991a) has shown that $c = 1$ minimises the risk of the $\hat{\sigma}_{LS}^2$ estimator. The risk minimising critical value for the $\hat{\sigma}_{ML}^2$ estimator varies according to whether a minimax or a minimin criterion is chosen with the optimal being $c = 0$ or $c = \infty$ respectively.⁹

Giles and Lieberman (1991b) apply Brook's (1976) criterion of minimax regret to the three classes of estimators to determine alternative optimal critical values and find that the minimax regret critical value lies in the ranges (1.4,7.2), (1.3,1.5) and (1.3,2.7) for the $\hat{\sigma}_{ML}^2$, $\hat{\sigma}_{LS}^2$ and the $\hat{\sigma}_{MS}^2$ estimators respectively. They conclude that the risk minimising critical values of $c = 0, 1$ and $\nu/(\nu+2)$ for $\hat{\sigma}_{ML}^2$, $\hat{\sigma}_{LS}^2$ and $\hat{\sigma}_{MS}^2$ respectively are easier to

⁹ Because the PTE may dominate the unrestricted estimator in the LS and MS cases the $\max_{\lambda}(\rho(\hat{\sigma}_i^2, \sigma^2)) = \rho(s_i^2, \sigma^2)$ because $\rho(\hat{\sigma}_i^2, \sigma^2) \rightarrow \rho(s_i^2, \sigma^2)$ from below as $\lambda \rightarrow \infty$, where $i = LS$ or MS . Therefore the minimax criterion is inappropriate in these cases.

apply and generally give rise to PTEs with lower risks than those which use the minimax regret critical values.

2.5 Pre-Test Estimation in a Mis-Specified Model

Although thus far we have only considered the literature pertaining to a correctly specified regression model, applied econometricians inevitably must have to deal with models containing some degree of mis-specification. Two examples of model mis-specification which may commonly occur in practice are; a mis-specification of the design matrix arising as a result of incorrect theory leading to relevant regressors being omitted from the model, or because the theoretically correct variables are unobservable and/or a mis-specification of the form of the error process. In recent years a number of authors have addressed the problem of applying the traditional restricted, unrestricted and pre-test estimators in a model framework that suffers from some type and degree of mis-specification.

The first such author, Ohtani (1983) considers a simple regression model in which a relevant regressor is unobservable and is proxied by another, less relevant but observable variable. The true data generating process is

$$y = x\beta + z\gamma + u; u \sim N(0, \sigma^2 I_T),$$

where y and x are $T \times 1$ vectors of observations, z is a $T \times 1$ vector of the unobservable regressor and u is the disturbance term. To test the hypothesis $H_0: \gamma = 0$ vs $H_A: \gamma \neq 0$ Ohtani proxies the z vector using a $T \times 1$ vector of observations of a variable p . the restricted and unrestricted models are now

$$y = x\xi + p\zeta + \omega; \quad \omega = x(\beta - \xi) + z\gamma - p\zeta + u$$

and

$$y = x\xi^* + u; \quad u = x(\beta - \xi) + z\gamma + u,$$

respectively. In his analysis Ohtani treats the ω and v disturbances as though they were well behaved $N(0, \sigma^2 I_T)$ random vectors. The statistic used to test the hypothesis is

$$F = \hat{\xi}^2 ((x'x)(p'p) - (x'p)^2) / (x'x) \hat{\sigma}^2,$$

where $\hat{\sigma}^2 = (y - x\hat{\xi} - p\hat{\zeta})'(y - x\hat{\xi} - p\hat{\zeta}) / (T-2)$, $\hat{\xi}$ and $\hat{\zeta}$ denote the OLS estimators of ξ and ζ respectively. This statistic is distributed as a doubly non-central $F'_{(1, T-2; \lambda_1, \lambda_2)}$ statistic with non-centrality parameters $\lambda_1 = \frac{1}{2}(t_\gamma^2 r_{zp.x}^2)$ and $\lambda_2 = \frac{1}{2}(t_\gamma^2 (1 - r_{zp.x}^2))$, where t_γ denotes the ratio of the true γ to the standard error of $\hat{\gamma}$ and $r_{zp.x}^2$ denotes the partial correlation coefficient for z and p given¹⁰ x . Ohtani derives the risks of the OLS, RLS and PT predictors and among other things, shows that the pre-test predictor may have the smallest risk of the three over some part of the parameter space. This is in contrast to the result that, in a correctly specified model, the PTE never has the minimum risk of the three estimators and will in fact have the greatest risk over some part of the parameter space.

In his appendix Ohtani generalises his simple model to one with K regressors, K_1 of which are observable and K_2 of which are not, and general restrictions. The general model is

$$y = X\beta + Z\gamma + u = \bar{X}\bar{\beta} + u; u \sim N(0, \sigma^2 I_T),$$

where y is a $T \times 1$ vector containing observations on the dependent variable, X is a $T \times K_1$ non-stochastic matrix of rank K_1 containing observations on K_1 regressor variables, Z is a non-stochastic $T \times K_2$ matrix of K_2 unobservable

¹⁰ Ohtani defines the non-centrality parameters as $\lambda_1 = t_\gamma^2 r_{zp.x}^2$ and $\lambda_2 = t_\gamma^2 (1 - r_{zp.x}^2)$ in the main body of his paper, however these definitions are not consistent with the formulae given in the appendix where they are derived analytically.

variables, $\bar{X} = [X, Z]$ and $\bar{\beta}' = [\beta', \gamma']$. The Z matrix is proxied by a $T \times K_2$ non-stochastic, observable, matrix P, which is of rank K_2 . The model including the proxy variable is

$$y = X\xi_1 + P\xi_2 + \omega = X_*\xi + \omega; \omega = X(\beta - \xi_1) + Z\gamma - P\xi_2 + u,$$

where $X_* = [X, P]$ and $\xi' = [\xi'_1, \xi'_2]$. The hypothesis to be tested is described by $H_0: H'\xi = h$ vs $H_A: H'\xi \neq h$, where H is a $K \times J$ non-stochastic matrix of rank J, and the restricted and unrestricted estimators are given by $d = S_*^{-1}X'_*y$ and $d^* = d - S_*^{-1}H(H'S_*^{-1}H)^{-1}(H'd - h)$ respectively, $S_* = X'_*X_*$. The test statistic is the usual Wald statistic given by $F = (H'd - h)'(H'S_*^{-1}H)^{-1}(H'd - h)/J\hat{\sigma}^2$ in this case, where $\hat{\sigma}^2 = (y - X_*d)'(y - X_*d)/(v)$. This statistic is distributed with a doubly non-central $F'_{(J, v; \lambda_1, \lambda_2)}$ distribution with $\lambda_1 = (H'S_*^{-1}X'_*\bar{H}\bar{\beta} - h)'(H'S_*^{-1}H)^{-1}(H'S_*^{-1}X'_*\bar{H}\bar{\beta} - h)/2\sigma^2$ and $\lambda_2 = \bar{\beta}'\bar{X}'(I_T - X_*S_*^{-1}X'_*)\bar{X}\bar{\beta}/2\sigma^2$. Following the approach of Toyoda (1976), Ohtani shows that the risk of the pre-test predictor, defined as

$$X\hat{\xi} = \begin{cases} X_*d; & \text{if } F \geq c \\ X_*d^*; & \text{if } F < c \end{cases}$$

is given by

$$\rho(X_*\hat{\xi}, E(y/X, Z)) = \sigma^2(\lambda_2 + K + (4\lambda_1 - J)h'(2, 0) - 2\lambda_2 h'(4, 0))$$

where

$$h'(i, j) = \Pr. \left(\frac{\chi'_{(j+1; \lambda_1)}}{\chi^2_{(v+j; \lambda_2)}} < \frac{cJ}{T-K} \right).$$

Note that one special case of Ohtani's model is where $P = 0$, that is the relevant regressors in the Z matrix have been completely excluded from the model specification. Unaware of Ohtani (1983), Mittelhammer (1984) considers the properties of the OLS, RLS, PTE and Stein rule predictors in a model

which has been mis-specified in the latter way. Mittelhammer's data generating process is given by

$$y = X_1\beta_1 + X_2\beta_2 + u = X\beta + u, \quad 2.5.1$$

where Y is a $T \times 1$ vector, $X = [X_1, X_2]$ is a non-stochastic $T \times K$ matrix of rank K which has been partitioned into two submatrices, X_1 and X_2 , each of full rank and sizes $T \times K_1$ and $T \times K_2$ respectively, $\beta' = [\beta'_1, \beta'_2]$ is a $T \times 1$ coefficient vector and the disturbance term u is $N(0, \sigma^2 I_T)$ distributed.

Mittelhammer specifies the fitted model as

$$y = X_1\beta_1 + \epsilon, \quad 2.5.2$$

where ϵ is regarded as being $N(0, \sigma^2 I_T)$ distributed, although the correct distribution is $\epsilon \sim N(\xi, \sigma^2 I_T)$, with $\xi = X_2\beta_2$. The hypothesis to be tested is written in the usual way as $H_0: R\beta = r$ vs $H_A: R\beta \neq r$ and the level of the hypothesis error is defined as¹¹ $\delta = R\beta - r$. The restrictions are tested using the usual Wald statistic given above in 2.3.1 which is distributed as a doubly non-central $F'_{(J, \nu; \lambda_1, \lambda_2)}$ random variable with $\lambda_1 = (RS^{-1}X'_1\xi + \delta)'(RS^{-1}R')^{-1}(RS^{-1}X'_1\xi + \delta)/2\sigma^2$ and $\lambda_2 = \xi'(I_T - X_1S_1^{-1}X'_1)\xi/2\sigma^2$, where $S_1 = X'_1X_1$. Mittelhammer shows that

$$\rho(X_1b, E(y)) = (K_1 + 2\lambda_2)\sigma^2,$$

$$\rho(X_1b^*, E(y)) = (K_1 + 2\lambda_2 + 2\lambda_1 - J)\sigma^2$$

and

$$\rho(X_1\hat{\beta}, E(y)) = (K_1 + 2\lambda_2 + 4\lambda_1 - Jh'(2, 0) - 2\lambda_1h'(4, 0))\sigma^2$$

where $b = S_1^{-1}X'_1y$, $b^* = b - S_1^{-1}R'(RS_1^{-1}R')^{-1}(Rb - r)$ and $\hat{\beta} = b^*I_{(0, c]}(u) + bI_{(c, \infty)}(u)$.

¹¹ Mittelhammer actually defines $\delta = r - R\beta$. However, to be consistent with our earlier notation, we will not follow his convention.

Mittelhammer notes that the RLS predictor is risk superior to the OLS predictor if $\lambda_1 \leq \frac{J}{2}$. This condition is the same as in the case where the model is well specified, except for the definition of λ_1 . However, Mittelhammer points out that, unless $X_1'X_2\beta_2 = 0$, λ_1 will be non-zero even if the restrictions are valid. That is, $\delta = 0$ does not necessarily imply that $\lambda_1 = 0$. In fact if $(RS^{-1}X_1'\xi)'(RS^{-1}R')^{-1}(RS^{-1}X_1'\xi)/2\sigma^2 \geq \frac{J}{2}$ the RLS predictor will be risk inferior to the OLS predictor in the case of the restrictions being valid. Therefore the imposition of valid restrictions serves to increase estimation risk in this case. Similarly, the PT predictor may be risk inferior to the OLS predictor even if the hypothesis being tested is valid.

Giles (1986) considers the opposite case to the one considered by Mittelhammer (1984), namely that of including irrelevant regressors in a regression model. Giles assumes that the true data generating process is given by 2.5.2 above and that the error term is well behaved with $\epsilon \sim N(0, \sigma^2 I_T)$ but that the fitted model is given by 2.5.1 above with the error term in this case being regarded as being well behaved, when in fact it is $N(-\xi, \sigma^2 I_T)$ distributed. The hypothesis under consideration is described by $H_0: R\beta = r$ vs $H_A: R\beta \neq r$ and is tested using the usual Wald test statistic given in 2.3.1. Giles finds that overfitting the model leaves the distribution of the test statistic and the risk functions of the estimators the same as in the case of a well specified model except for a scaling of the non-centrality parameter. Giles notes that if the restrictions involve only the relevant regressors then the usual results hold exactly.

The PTE of the regression scale parameter in a model mis-specified through the exclusion of relevant regressors is considered by Giles and Clarke (1989). The basic model used is that of Mittelhammer (1984) described in 2.5.1 above, and they consider the Maximum Likelihood s_{ML}^2 , s_{ML}^{*2} and $\hat{\sigma}_{ML}^2$

estimators as defined above. They show that the risks of these estimators are

$$\rho(s_{ML}^2, \sigma^2) = \sigma^4 (2(\nu_1 + 4\lambda_2) + (2\lambda_2 - K_1)^2) / T^2,$$

$$\rho(s_{ML}^{*2}, \sigma^2) = \sigma^4 (2[J + \nu_1 + 4(\lambda_1 + \lambda_2)] + [J - K_1 + 2(\lambda_1 + \lambda_2)^2]) / T^2,$$

and

$$\begin{aligned} \rho(\hat{\sigma}_{ML}^2, \sigma^2) = & \rho(s_{ML}^2, \sigma^2) + \sigma^4 (2J\nu_1 h'(2,2) - 2JTh'(2,0) \\ & + 4\lambda_1 \nu_1 h'(4,2) + [J(J+2) - 4\lambda_1 T] h'(4,0) \\ & + 4(J+2)\lambda_1 h'(6,0) + 4\lambda_1^2 h'(8,0) \\ & + 4\lambda_2 [Jh'(2,4) + 2\lambda_1 h'(4,4)]) / T^2. \end{aligned}$$

As in the well specified model, the risks depend only on T , J , K , and the non-centrality parameters. Giles and Clarke evaluate these functions numerically and show that the effects of the model mis-specification on the three estimators are qualitatively the same as the effects of this mis-specification on the OLS, RLS and PT predictors as reported by Mittelhammer (1984). In particular it is apparent that, if the mis-specification is severe enough, the unrestricted estimator may dominate both the restricted and pre-test estimators even if the restrictions are in fact valid. Giles and Clark also note that the restricted and PT estimators may be strictly dominated by the unrestricted estimator if λ_2 is sufficiently large. This is in contrast to Mittlehammer's result that the unrestricted predictor will have lower risk than the unrestricted predictor when $\lambda_1 < \frac{J}{2}$ regardless of the value of λ_2 .

One question that is raised by the application of the PTE in a mis-specified model is that of the appropriateness of the optimal critical value rules of Toyoda and Wallace (1976) or Brook (1976). This question is addressed by Giles, Lieberman and Giles (1992) who determine the optimal

critical value for the pre-test using Brook's minimax regret criterion in a model mis-specified by the exclusion of relevant regressors. The model used is the same as Mittelhammer's (1984) model given in 2.5.1 above and the predictors under consideration are the classical OLS, the RLS and PT predictors. They replicate Brook's (1976) results in the case where $\lambda_2 = 0$, as this implies that either, no mis-specification is present, or that the excluded regressors are orthogonal to the included regressors. In this case, therefore, the predictors are not affected by the mis-specification. The optimal critical value is approximately two regardless of the degrees of freedom of the pre-test. However this is not the case if λ_2 differs from 0. Not only does the optimal critical value differ from two, but it is no longer approximately invariant as the degrees of freedom change. They also note that, for given degrees of freedom, the optimal critical value decreases as the degree of model mis-specification increases. This implies that, other things being equal, the optimal pre-test size is increasing in λ_2 , emphasising the point made by both Toyoda and Wallace (1976) and Brook (1972, 1976) that the optimal pre-test size is generally greater than the traditional value of 5%.

Another form of mis-specification is considered by Giles (1991a & b), namely that of a mis-specification of the error distribution. In Giles (1991a) the problem is one in which the disturbance term is spherically symmetrically distributed, $\epsilon \sim \text{SSD}_N(0, I_T)$, as opposed to being normally distributed.¹² The Probability Density Function, pdf, of ϵ is given by

$$f(\epsilon) = \int_0^{\infty} f_N(\epsilon)f(\tau)d\tau, \quad 2.5.3$$

¹² The normal distribution is a member of the spherically symmetric family.

where $f_N(\varepsilon)$ is the normal pdf and $f(\tau)$ is the pdf of τ which is supported on $[0, \infty)$. Giles assumes that the non-normality of the error term is not taken into account in the fitting the model to the data. In other respects the problem is well specified and is described by 2.2.1 above. The hypothesis to be tested is described in the usual way as $H_0: R\beta = r$ vs. $H_A: R\beta \neq r$ and is tested using the usual Wald statistic given in 2.3.1.

Although the model is mis-specified in terms of the form of the error distribution, Box (1952) shows that, under the null hypothesis, the statistic is distributed as a central $F_{(J, \nu)}$ statistic as is the case in a well specified model. The distribution of the statistic under the alternative hypothesis, is, however, dependent on the particular type of spherically symmetric process that the errors follow. This fact notwithstanding, King (1979) shows that the Wald statistic provides a UMPI size α test for the restrictions as long as the errors follow an elliptically symmetrical distribution.

Giles derives the distribution of the test statistic and the restricted, unrestricted and pre-test predictor risks under the assumption of spherically symmetric errors and in the case of multivariate t distributed errors.¹³

Giles notes from these formulae that the restricted estimator dominates the unrestricted estimator over a larger part of the parameter space than would be the case if the error term was normally distributed. That is, the value of λ at which the restricted and the unrestricted estimators' risks are equal is greater than $\frac{J}{2}$, the latter being the value at which the risks equate if the errors are normally distributed. From her numerical evaluations of the risk functions in the case of multivariate t errors Giles notes that

¹³ In the latter case, $f(\tau)$ is an inverted gamma distribution with scale parameter denoted σ^2 and degrees of freedom denoted μ . The error distribution collapses to the normal distribution if $\mu = \infty$.

qualitatively the estimators' risk functions have the same characteristics as they have in the well specified case. She suggests that the difference between an estimator's risk under normality and multivariate t errors is not significant for¹⁴ $\mu \geq 100$.

Giles then considers the estimators of the scale parameter, $\sigma_e^2 = E(\tau^2)$. The unrestricted, restricted and pre-test estimators are $s_e^2 = (y-Xb)'(y-Xb)/\nu$, $s_e^{*2} = (y-Xb^*)'(y-Xb^*)/(\nu+J)$ and $\hat{\sigma}_e^2 = s_e^{*2}I_{(0,c)}(u) + s_e^2I_{(c,\infty)}(u)$ respectively.

Giles derives the risk of these estimators under the assumption of spherically symmetric disturbances and in the special case of multivariate t distributed disturbances. From her numerical evaluations of these formulae, She notes that the PTE risk functions change in a similar way to the risk functions of the prediction vector estimators as the normality assumption is relaxed. Estimator risks increase and the PTE converges to the unrestricted estimator at a slower rate as μ decreases. Also, the PTEs which correspond to a choice of critical value of $c \in (0,1]$ dominate the unrestricted estimator over the parameter space, and dominate the unrestricted estimator over part of the parameter space. The PTE which corresponds to a critical value of $c = 1$ dominates the other members of this family.

Giles also notes that for some values of μ the PTE may dominate the restricted estimator, depending on the choice of critical value and the degrees of freedom of the pre-test. If $\mu \leq 15$, the restricted estimator is

¹⁴ In a later paper Wong and Giles (1991) note that there are some instances in which the PT predictor dominates both of its component over a region of the parameter space. This is in contrast to the usual result, noted above, that in a well specified model pre-testing is never the preferred strategy, given perfect information. Wong and Giles find that it is the values of J and μ , but not ν , which determine whether or not such a region, in which the PTE dominates, exists and how large it is. In their numerical evaluations they find that the region increases in size as J increases and as μ decreases.

generally dominated by the PTE for all critical values except c close to zero or c very large.

Giles (1991b) extends this problem further to a case in which the error distribution is mis-specified as being Normal when it is, in fact, spherically symmetric and the regressor matrix is mis-specified by the exclusion of relevant regressors. The basic model is described by 2.5.1 above with the error process given by 2.5.3 above. The test and estimators are the same as those considered by Giles (1991a).

Giles derives the risk functions of the unrestricted, restricted and pre-test predictors in the case of spherically symmetric errors and in the special case of Multivariate t distributed errors. She notes that the results described by Mittelhammer (1984) which relate to a model with Normally distributed errors and a mis-specified regressor matrix, are qualitatively the same in the case of spherically symmetric errors.

Giles also considers the estimators of the scale parameter σ_e^2 as defined above. She derives the risk functions of the restricted, unrestricted and pre-test estimators in the case of spherically symmetric errors and in the special case of Multivariate t distributed errors. She notes that, qualitatively, many of the characteristics of these functions are the same as those in the model discussed by Giles and Clarke (1989) who consider maximum likelihood estimation of σ^2 in a model where the errors are normally distributed and the regressor matrix is mis-specified by the exclusion of relevant regressors. For example, the risks of all three estimators increase without bound as λ_2 increases, as we would expect, while the unrestricted estimator risk is independent of λ_1 and the PTE risk is bounded for given λ_2 .

Giles' numerical evaluations suggest that for all feasible μ it is generally better to pre-test using $c = 1$ rather than to use the restricted estimator, even in the case of valid restrictions. This contrasts with Giles

(1991a), who found that the restricted estimator generally dominates the PTE in the neighbourhood of the null when $\mu > 15$.

2.6 Concluding Remarks

In this chapter we have reviewed the literature which deals with the finite sample properties of the estimators of the coefficient vector, prediction vector and scale parameter after a pre-test of exact linear restrictions on the coefficient vector in the classical linear model. There are, however, many other pre-test problems considered in the econometrics literature. Outside the context of regression, there are, for example, a number of papers which consider the estimation of the scale parameter under normal sampling, after a pre-test for homoscedasticity. This problem, termed the "pooling problem" was first considered by Bancroft (1944) who considers the bias and variance of the estimator of the scale parameter after such a pre-test. Bancroft shows that for an appropriate critical value of the pre-test, the PTE strictly dominates the unrestricted "never pool" estimator. Toyoda and Wallace (1975), Ohtani and Toyoda (1978) also consider this PTE and determine the optimal critical value of the pre-test under a criterion of minimizing the average relative risk, where the pre-test has a one-sided alternative, and under a criterion of minimax regret respectively. See also Giles (1992a), who derives the exact distribution of the PTE considered by Bancroft, and Giles (1992b) who considers the problem in the context of linear regression with spherically symmetric disturbances.

The related problem, that of estimating the coefficient vector after a pre-test for homoscedasticity, is considered by Greenberg (1980). In this case the unrestricted estimator is the Two Stage Aitken Estimator 2SAE, the finite sample properties of which are derived by Taylor (1978). See also

Mandy (1984) and Ohtani and Toyoda (1980) who consider the problem in terms of a model with orthonormal regressors, and Yancey *et al.* (1984).

A similar problem is that of estimating the coefficient vector after a preliminary test for autocorrelation and a number of studies have considered this issue. Papers in this area include Judge and Bock (1978), Fomby and Guilkey (1978), Griffiths and Beesley (1984), King and Giles (1984) and Giles and Beattie (1987). One of the recommendations to come out of these studies is that the optimal level of significance of the pre-test may be larger than the traditional value of $\alpha = 5\%$. Fomby and Guilkey recommend a value of 50%. This is supported by King and Giles and, more recently, Giles and Lieberman (1992a) who consider the problem of testing for autocorrelation after the application of a preliminary t-test.

This issue of pre-test testing, that is the effect of a preliminary test on the properties of subsequent tests applied to the model, is also considered by Nakamura and Nakamura (1978) and King and Giles (1984) in the context of testing for linear restrictions after a preliminary test for autocorrelation. Other references include Gurland and McCullough (1962), Ohtani and Toyoda (1985a), Toyoda and Ohtani (1986) and Ohtani and Toyoda (1988) who consider the application of a test for heteroscedasticity in the errors prior to testing for linear restrictions on the coefficients of the regression model and Ohtani (1988) who considers the opposite problem, namely that of testing for linear restrictions prior to testing for heteroscedasticity in the errors. These studies, like those of Giles and Lieberman (1992a) and King and Giles (1984) emphasize the importance of the chosen pre-test size on the model and recommend sizes substantially larger than those traditionally used in applying pre-tests. Other pre-test situations which have received attention in the literature include:

Estimation after a pre-test of inequality, rather than exact equality constraints. See, for example, Judge and Bock (1983), Judge and Yancey (1981, 1986), Yancey *et al.* (1989) and Wan (1992).

Estimation after a preliminary Hausman (1978) test for specification error. See Morey (1984) and Gouriéroux and Trognan (1984).

Estimation of a seemingly unrelated regression model using either OLS or generalised least squares depending on the outcome of a pre-test for correlation between the errors of the equations. See Ozcam *et al.* (1991).

Estimation after a pre-test for ARCH(1) errors. See Engle *et al.* (1985).

Estimation after multiple pre-tests. See Ozcam and Judge (1991) who consider estimation of the coefficient vector in a model with orthonormal regressors after prior tests for heteroscedasticity of the disturbance followed by a test of linear restrictions.

For further discussion of these and further examples of pre-test estimators see Giles and Giles (1993).

We have assumed throughout this chapter that, unless a non-scalar error covariance matrix is the subject of the pre-test, the regression disturbances are independently and identically distributed. However there are a number of reasons why the assumption of a scalar error covariance matrix might be invalid. We consider these in the next chapter.

CHAPTER THREE

THE EFFECT OF A MIS-SPECIFIED ERROR COVARIANCE MATRIX ON A PRE-TEST FOR LINEAR RESTRICTIONS

3.1 Introduction

The effects of a mis-specification of the error covariance matrix on particular tests of restrictions on the coefficients of the linear regression model have been considered by a number of authors. Such a mis-specification has been shown to have varying effects, depending on the true form of the error covariance matrix and the form of the restrictions as well as on the particular test statistic employed. In terms of testing exact linear restrictions, the most commonly used test is based on the Wald statistic described by 2.3.1 above. In a well-specified model the Wald statistic provides an UMPI size α -test of exact linear restrictions. However, it is well known that the presence of a non-scalar error covariance matrix can distort both the size and power of this test. Various asymptotically valid alternative tests which are designed to outperform the Wald test in the presence of heteroscedasticity or autocorrelation have been proposed; see, for example, Watt (1979), Jayatissa (1977), or Weerahandi (1987) among others. However, such statistics may have very poor small sample properties, as is pointed out by Honda (1982), Ohtani and Toyoda (1985b) and Griffiths and Judge (1990). The Wald test remains the most commonly applied test and it is the one that we consider here.

The Wald test statistic also has the advantage that it can be written as a ratio of quadratic forms in a normal random vector thereby making the calculation of its exact size and power possible using the algorithms of

Imhof (1961) or Davies (1980), for example. It is this property which we exploit in this thesis when deriving the risks of the pre-test estimators which result from the application of this test.

In this chapter we briefly survey the literature which deals with the forms of non-scalar error covariance matrix that may arise in an applied situation. As it is our intention to consider the consequence of incorrectly assuming a scalar error covariance matrix, we do not attempt to survey the tests proposed to detect such a problem.

In an applied situation such tests may well be carried out after pre-tests relating to the specification of a regression model and therefore it is commonly the case that such pre-tests are carried out in a model in which a possibly non-scalar error covariance matrix has not been allowed for. Section 3.2 surveys the causes of several heteroscedastic processes which arise in applied econometric situations and the effect of one of these on a particular pre-test for linear restrictions while section 3.3 considers the causes and effects of an uncorrected autoregressive process in the errors.

3.2 Heteroscedastic Errors in the Linear Regression Model

In the general linear model, heteroscedasticity exists if the error variance is changing over the sample period; *i.e.*, if the diagonal elements of the error covariance matrix are not all identical. The usual assumption of a constant error variance is likely to be an unrealistic one when using regressors based on average data as, if the data used in estimation are derived by averaging observations within subsamples or periods, the variance of each of the data values is inversely proportional to the size of the

subsample from which it comes,¹ or when using cross-sectional data. Heteroscedasticity may also occur if there are one or more structural changes of some kind in the data generating process. In this case, the error variance may be constant within subgroups of observations but not necessarily over all observations in the total sample period. In this case the regression model can be written

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_a \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_a \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_a \end{bmatrix}$$

where each y_i is a $T_i \times 1$ vector, each X_i a $T_i \times K$ non-stochastic matrix and each ε_i is a $T_i \times 1$ disturbance term with $\varepsilon_i \sim N(0, \sigma_i^2 I_{T_i})$, $i = 1, \dots, a$.

An alternative model of heteroscedasticity is one in which the error variance is some function of a set of exogenous variables such as the regressor variables. Consider, for example, the random coefficient model of Hildreth and Houck (1968). In this case the linear model may be written

$$y_t = \sum_{k=1}^K (\beta_k + v_{tk}) x_{tk} = \sum_{k=1}^K \beta_k x_{tk} + \varepsilon_t \quad ; \quad t = 1, \dots, T,$$

where v_{tk} represents the stochastic element of the regression coefficient with each v_{tk} i.i.d. with zero mean and $E(v_{tk}^2) = \alpha_k$. The error variance, then, is a linear function of a vector of non-stochastic parameters $z_t = (1, z_{t1}, \dots, z_{tn})$ such that $E(\varepsilon_t^2) = \sigma_t^2 = z_t' \alpha$. In this case each $z_{tk} = x_{tk}^2$ and $n = K$. In general the values of z_t may not be related to the regressors. See also Amemiya (1977) and Froehlich (1973) among others.

¹ See Judge *et al.* (1985, p.419) for an example of this.

A similar heteroscedastic process that has received attention in the literature is one in which the disturbance variance is related to one or more of the regressors raised to a particular power. Consider, for example, the case of estimating household expenditure functions. In this context it has been suggested that the variability of the dependent variable, household expenditure, is inversely related to household income, see, for example, Battese and Bonyhady (1981) who consider the model suggested by Prais and Aitchison (1954) in which the variance of the dependent variable is proportional to an unknown power of its expectation. Therefore, in the framework of the general linear model, we have $E(\varepsilon_t^2) = \sigma_t^2 = \sigma^2(x_t'\beta)^p$. It is commonly assumed in the literature relating to this process that $p = 2$.

The more general concept of multiplicative heteroscedasticity is suggested by Harvey (1976). In this case σ_t^2 is assumed to be related to some, non-stochastic, $n \times 1$ vector z_t such that $\sigma_t^2 = \exp(z_t'\alpha)$. The first element of the vector z_t is assumed to be a constant. If we consider a special case, in which $\alpha' = [\log_e \sigma^2, p]$ and $z_t' = [1, \log_e x_{tk}]$, this specification can be written $\sigma_{tk}^2 = \sigma^2 x_{tk}^p$ where the variance is proportional to some power of one of the explanatory variables. This approach then subsumes that of Park (1966) and Geary (1966) who consider this particular process.

A number of tests for detecting heteroscedasticity have been proposed, such as the Breusch and Pagan (1979) Lagrange multiplier test. Breusch and Pagan assume that the error variance in the model

$$y_t = x_t'\beta + \varepsilon_t \quad ; \quad t = 1, \dots, T$$

where ε_t is normally and independently distributed with mean 0 and variance σ_t^2 , is related to some non-stochastic $n \times 1$ vector z_t (the first element of which is unity) by the relationship $\sigma_t^2 = h(z_t'\alpha)$ for some twice differentiable

function $h(\cdot)$. The null hypothesis, that of homoscedasticity, may be written $H_0: \alpha_2 = \alpha_2 = \dots \alpha_n = 0$. Breusch and Pagan define $s^2 = \frac{1}{T}(y-Xb)(y-Xb)$, where b is the OLS estimator of the regression coefficient vector, and show that the test statistic, defined as one half the explained sum of squares from the regression of $s^{-2}(y_t - x_t' b)^2$ on z_t is asymptotically distributed as $\chi^2_{(n-1)}$ under the null hypothesis. Other tests have been proposed by Bartlett (1937) (see also Dyer and Keating (1980) who consider the determination of critical values for this test), Goldfeld and Quandt (1965), Szroeter (1978) and White (1980), among others. See Judge *et al.* (1985) for further discussion and references.

As noted above, such tests may well be carried out, in an applied situation, after tests relating to the specification of the model. Therefore we now turn our attention to the effect of uncorrected heteroscedasticity on such tests, in particular a test for linear restrictions on the coefficients of this model.

3.2.1 The Effects of Uncorrected Heteroscedasticity on the Chow Test

The effect of heteroscedasticity on a pre-test for linear restrictions was first considered by Toyoda (1974) in terms of applying a Chow (1960) test for structural change in the model

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad 3.2.1$$

where y_1 and X_1 are $T_1 \times 1$ and $T_1 \times K$, with X_1 non-stochastic and of rank K . The β_1 are $K \times 1$ coefficient vectors and the ε_1 are $N(0, \sigma_1^2 I_{T_1})$ random variables.

The error covariance matrix of ε is

$$\begin{bmatrix} \sigma_1^2 I_{T_1} & 0 \\ 0 & \sigma_2^2 I_{T_2} \end{bmatrix} = \sigma^2 \Omega, \text{ where } \Omega = \begin{bmatrix} I_{T_1} & 0 \\ 0 & \psi & I_{T_2} \end{bmatrix}$$

and $T_1 + T_2 = T$. If, based on the outcome of the Chow test, the hypothesis that there is no structural change is accepted, the model is rewritten as

$$y \equiv \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} \epsilon'_1 \\ \epsilon'_2 \end{pmatrix} \equiv X\beta + \epsilon.$$

The restrictions to be tested in this case are represented by $H_0: \beta_1 = \beta_2$ vs $H_A: \beta_1 \neq \beta_2$. The Wald statistic can be written as

$$u = \frac{(e'e - (e'_1 e_1 + e'_2 e_2)) / K}{(e'_1 e_1 + e'_2 e_2) / (T - 2K)}$$

where $e = y - Xb$ and $e_i = y_i - X_i b_i$ with $b = (X'X)^{-1}X'y$ and $b_i = (X'_i X_i)^{-1}X'_i y_i$, $i = 1, 2$. In this case b represents the restricted estimator, as it is applied when the restriction, that there has been no structural change, is accepted, while b_1 and b_2 represent the unrestricted estimators of β_1 and β_2 respectively.

Toyoda considers the robustness of the Chow test to a heteroscedastic error process in this situation by showing that u is approximately $\frac{T-2K}{f}$ $F_{(K,f)}$ distributed, where

$$f = \frac{\left[(T_1 - K)\sigma_1^2 + (T_2 - K)\sigma_2^2 \right]^2}{(T_1 - K)\sigma_1^4 + (T_2 - K)\sigma_2^4}.$$

Therefore the approximate size of the test is given by

$$\Pr.(u > c) \approx \Pr.\left[F_{(K,f)} > \frac{f}{T-2K} c\right] = \alpha^*,$$

where c is chosen such that $\Pr.\left[F_{(K,T-2K)} > c | H_0 \text{ true}\right] = \alpha\%$, the nominal size of the test. Toyoda notes that α^* will be at a minimum when $\sigma_1^2 = \sigma_2^2$, therefore his approximation suggests that the presence of heteroscedasticity

in the disturbance term is likely to bias the size of the test upward. However, other things being equal, if at least one of the samples is of large size, the difference between the nominal and approximate real significance levels is minimal. As T_1 and/or $T_2 \rightarrow \infty$, $\alpha^* \rightarrow \alpha$ for all values of $\psi = \sigma_1^2/\sigma_2^2$.

Toyoda evaluates the approximate real test size for a variety of degrees of heteroscedasticity with ψ ranging from $\psi = 0.0$ to $\psi = 100$, and various values of T_1 , T_2 and K . He notes that the discrepancy between α and α^* is greater the greater is the degree of heteroscedasticity and the greater is the number of regressors in the model. There are, however, some problems with the approximation used by Toyoda, as is noted by Schmidt and Sickles (1977).

Schmidt and Sickles consider the same problem as Toyoda but note that the test statistic, u , can be written as a ratio of quadratic forms in the normal random vector e . They write $u = (T_1 + T_2 - 2K)/K \left(\frac{e' Ae}{e' Be} \right)$ where $B = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}$, $A = M - B$ and $M_1 = I_{T_1} - X_1 S_1^{-1} X_1'$. The density function of u can be calculated using the method of Imhof (1961) or Davies (1980) and therefore the exact value of the test size can also be evaluated.

Schmidt and Sickles carry out this evaluation for a variety of artificial data sets in the context of a simple regression model. The data sets they use are

- i) each X_1 contains a constant and a linear trend.
- ii) each X_1 contains a constant and a vector of iid $N(0,1)$ random variables.
- iii) each X_1 contains a constant and a vector of $N(0,1)$ random variables with autocorrelation.

Having evaluated the exact pre-test size for a number of values of T_1 and K , Schmidt and Sickles conclude that Toyoda's (1974) approximation is not, in fact, very accurate. They note that the type of data used in forming the X_1

matrices may make a difference to the bias in the pre-test size and that, in general, the exact size is less than that given by Toyoda's approximation. In fact, the true size of the pre-test may be less than the nominal size. This contrasts with Toyoda's conclusion that the direction of the bias in the test size is always upward. Contrary to Toyoda's conclusion that the test is robust if the size of at least one sample is sufficiently large, Schmidt and Sickles also note that increasing the sample size does not increase the reliability of the pre-test, in fact the opposite may be true.

Unfortunately, the data dependency of the bias in the pre-test size implies that there is no general prescription which can be applied to offset the effects of the heteroscedasticity. Notwithstanding this, the Chow test continues to be used widely. Giles and Lieberman (1991a) address the problem of determining data independent bounds on the effect of heteroscedasticity on the test. Using the results of Kiviet (1980) they determine bounds on the true critical value for the test in the model described by 3.2.1 above in the following way. Under the null hypothesis $u_L \leq u \leq u_U$ where

$$u_L = \frac{T-2K}{K} \left(\frac{\sum_{i=1}^K \lambda_i \chi_i^2}{\sum_{i=2K+1}^T \lambda_i \chi_i^2} \right)$$

and

$$u_U = \frac{T-2K}{K} \left(\frac{\sum_{i=1}^K \lambda_{(T-K+1)} \chi_i^2}{\sum_{i=2K+1}^K \lambda_{(1-2K)} \chi_i^2} \right)$$

The χ_i^2 are independent central chi-squared variables with one degree of freedom and $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_T$ are the appropriately ordered diagonal elements of $\sigma^2 \Omega$, the error covariance matrix.²

² In fact, the results are the same if the eigenvalues of Ω rather than $\sigma^2 \Omega$ are used as the scale parameter cancels out of the formulae.

Giles and Lieberman calculate the bounds on the critical value, for a nominal test size of $\alpha\%$, as $c_L(\alpha)$ and $c_U(\alpha)$ where $\Pr.\left(u_L \geq c_L(\alpha)\right) = \alpha$ and $\Pr.\left(u_U > c_U(\alpha)\right) = \alpha$. They also calculate the bounds on the test size, α_U and α_L where $\Pr.(u_L \geq c) = \alpha_L$ and $\Pr.(u_U \geq c) = \alpha_U$, c is the critical value used for the test in the case where there is no heteroscedasticity, i.e., $\Pr.\left(F_{(K,T-2K)} \geq c\right) = \alpha$.

The bounds on the critical value and real size are independent of the data and depend only on T_1 , T_2 , K , α and $\psi = \sigma_2^2/\sigma_1^2$. Giles and Lieberman calculate these bounds for a variety of values of T_1 , T_2 , K and ψ for a nominal significance level of $\alpha = 5\%$. They note that, for $\psi < 1$, increasing T_1 leads to decreases in $c_L(0.05)$ and $c_U(0.05)$, and the converse occurs if $\psi > 1$. This pattern is generally reversed for increases in T_2 . Their results also show that the real pre-test size may vary substantially from α , with α_U taking values of up to 65.8% and α_L taking values as low as 0%, for the examples that they consider.

3.3 Autocorrelated Errors in the Linear Regression Model

Another form of error covariance matrix arises if the errors are correlated through time. In this case the error covariance matrix is not diagonal. For example, the case of a simple autoregressive process of order n is described by³ $\varepsilon_t = \sum_{i=1}^n \rho_i \varepsilon_{t-i} + \mu_t$; where ε_t is the value of the error at time t and $\mu_t \sim \text{iid}(0, \sigma^2)$. Such a process is denoted AR(n). Similarly, a

³ For stationarity we require that the roots of $1 - \rho_1 z - \rho_2 z^2 - \dots - \rho_n z^n$ to lie outside the unit circle. If $n=1$ this implies that $|\rho_1| > 1$.

moving average process of order n , MA(n) process is described by⁴ $\epsilon_t = \sum_{i=1}^n \bar{\rho}_i \mu_{t-i} + \mu_t$. Such a process may arise as a result of uncorrected excluding autocorrelated regressors from a regression model. The most commonly assumed process in both theoretical and empirical work is the AR(1) process and numerous tests have been devised to determine if such a process is present in regression disturbances, the most common being the test of Durbin and Watson (1950, 1951). Other tests include those of Durbin (1970a,b), Berenblut and Webb (1973) and King (1981) among others. There are also instances in which a higher degree autoregressive process is appropriate for a linear model. Thomas and Wallis (1971), for example, argue that, even in a model estimated using seasonally adjusted data, the error term may still suffer from fourth order correlation. Such a process may arise as a result of relevant regressors which have seasonal components being omitted from the model specification, measurement errors, or arbitrary human behaviour for example. Thomas and Wallis propose a simple test for an AR(4) process and, using a numerical example, demonstrate that such a process may be present in a seasonally adjusted model. They find that in their example, the presence of an AR(4) process causes the standard errors of the coefficients to be underestimated.

In a later paper, Wallis (1972) proposes a further test for fourth order autocorrelation. This test is based on the approach that Durbin and Watson (1950, 1951) used to construct their test for first order autocorrelation. The test is a bounds test and tables of significance points are given by

⁴ For stationarity we require $E[\epsilon_t^2] = \sigma_\mu^2(1 + \bar{\rho}_1^2 + \bar{\rho}_2^2 + \dots + \bar{\rho}_n^2)$ where σ_μ^2 is the variance of μ .

Wallis.⁵ Wallis applies his test in four quarterly economic models, those of Thomas and Wallis (1971), Nadiri (1969), Lipsey and Parkin (1970) and Mills (1962), and finds significant evidence of an AR(4) process in each case.

Other tests for AR(4) have been proposed, such as by King (1984, 1989) for example. A number of the tests for both AR(1) and AR(4) processes are reviewed by King (1987).

Another form of error generating mechanism which has been considered in a number of papers is the MA(1) process. For example, Pesaran (1973) and Balestra (1980) consider the estimation of regression models in which the errors follow such a process. It is possible to test for this using the Durbin Watson test, as a significant DW statistic can result from either AR(1) or MA(1) disturbances. King (1983b) points out that, in fact, the Durbin Watson test is approximately locally most powerful invariant test of uncorrelated errors against both AR(1) and MA(1) alternatives. Based on Monte Carlo evidence, Griffiths and Beesley (1984) suggest that if the errors are MA(1) distributed, with $\varepsilon_t = \rho_1 \mu_{t-1} + \mu_t$, applying a correction for AR(1) rather than MA(1) will nevertheless improve the efficiency of the parameter estimates. The question of differentiating between an AR(1) and an MA(1) process, given that there is evidence of some autocorrelation in the errors, has been addressed by a number of authors, including Walker (1967), King (1983a) and King and McAleer (1987) among others.

⁵ Extensions to these tables are given by Giles and King (1978), and the case of negative fourth-order autocorrelation is discussed by King and Giles (1977).

3.3.1 The Effects of Uncorrected Autocorrelation on a Pre-Test for Linear Restrictions

The effects of uncorrected autocorrelation on a pre-test are examined by Watson and Hannan (1956), Vinod (1976) and, more recently, by Kiviet (1980). Kiviet considers the general linear model

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \Omega)$$

with the hypothesis to be tested represented by $H_0: R\beta = r$ vs $H_A: R\beta \neq r$.

The test statistic is the usual Wald statistic

$$u = \frac{T-K}{J} \frac{(Rb-r)' (RS^{-1}R')^{-1} (Rb-r)}{(y-Xb)'(y-Xb)}.$$

Kiviet proves the result, noted above, that $u_L \leq u \leq u_U$ where

$$u_L = \frac{T-K}{J} \left(\frac{\sum_{i=1}^J \lambda_i \chi_i^2}{\sum_{i=K+1}^T \lambda_i \chi_i^2} \right)$$

$$u_U = \frac{T-K}{J} \left(\frac{\sum_{i=1}^J \lambda_{(T-J+1)} \chi_i^2}{\sum_{i=K+1}^T \lambda_{(1-K)} \chi_i^2} \right),$$

where χ_i^2 are independent chi square random variables with one degree of freedom and $\lambda_1 \leq \dots \leq \lambda_n$ are the eigenvalues of the error covariance matrix $\sigma^2 \Omega$. Kiviet evaluates the bounds on the true size and critical value for the tests assuming AR(1) and MA(1) and ARMA(1,1) errors.⁶

⁶ ARMA(1,1) errors can arise when the estimated model is the result of a transformation of another model in which the errors are generated by an AR(1) process. The ARMA(1,1) process can be represented by

$$\varepsilon_t - \rho \varepsilon_{t-1} = \mu_t + \bar{\rho} \mu_{t-1}$$

where ρ and $\bar{\rho}$ are the coefficients of the ARMA(1,1) process $|\rho| < 1$ and $|\bar{\rho}| < 1$ and $\mu_t \sim \text{iid } N(0, \sigma^2)$.

The upper bounds on the pre-test size, α_U , are data independent but vary with ρ , $\bar{\rho}$, J, T and K and range from 5% to 99% in the case of AR(1) errors, from 5% to 44% in the case of MA(1) errors and from 5% to 99% in the case of ARMA(1,1) errors in the examples studied by Kiviet.

Kiviet notes that the effects of increased sample size are moderate but that α_U increases with J, K and ρ or with $\bar{\rho}$. The effects of the MA(1) process on the size, while still significant, are less than the effects of the AR(1) process.

The lower bound on the test's size was found to be $\alpha_L \leq 1\%$ for all $\rho \geq .3$ in the case of AR(1) errors and $\alpha_L \leq 2\%$ for $\bar{\rho} \geq .3$ in the case of MA(1) errors. Both the AR(1) and MA(1) processes are special cases of the ARMA(1,1) process and the bounds on the true size of the test in this case depend on both ρ and $\bar{\rho}$ as well as on J, K and T.

A specific case in which a test of this type is carried out in the context of a mis-specified model is considered by Consiglieri (1987). The test in question is a Chow test for structural change and it is applied in the context of a model in which the errors terms are not mutually independent. The regression model is the same as the one considered by Toyoda (1974) and Schmidt and Sickles (1977) as in equation 3.2.1 above with Ω in this case a general positive definite matrix. Following the method of Koerts and Abrahamse (1969), Consiglieri shows that the cumulative distribution function of the test statistic can be expressed in terms of a weighted sum of independent χ_1^2 random variables and the true size of the test can therefore be calculated using the method of Imhof (1961), L'Esperance *et al.* (1976) or Davies (1980).

In her numerical evaluations, Consiglieri uses the same form of data as Schmidt and Sickles (1977); namely a constant regressor with either a linear

trend, an exponential trend, or a normally distributed (possibly autocorrelated) regressor. Consiglieri considers the effects of AR(1) errors on the true size of the test and concludes that the Chow test is very sensitive to a mis-specification of this kind. The true size of the pre-test can be as much as six times the nominal size when a trended regressor is used. Generally, the true size is greater (less) than the nominal size if the autocorrelation is positive (negative) and the larger the absolute value of ρ the worse is the distortion in test size. Only if $|\rho| < .25$ does she find that the test is fairly reliable.

An interesting feature of Consiglieri's results is that she finds increasing bias in the pre-test size as the total sample size increases, although the distortion in the test's size is less, the greater is the difference between the sizes of the two subsamples. See also Corsi *et al.* (1982) and Krämer (1989) in connection with this problem.

In a more recent paper, Giles and Scott (1992) consider the same model in terms of both AR(1) and MA(1) errors, and using a greater variety of data sets including orthonormal and uniformly distributed regressors, as well as real data, as regressors. Giles and Scott's results with AR(1) errors generally accord with the results of the earlier studies of Corsi *et al.* (1982) and Consiglieri (1981). The one exception they find is that, in the case of an orthonormal regressor set, the true size of the Chow test exceeds the nominal size of the test regardless of the sign of ρ . Giles and Scott find that the test is more robust when the errors follow an MA(1) process than it is in the case of AR(1) errors. Qualitatively, the effects of MA(1) errors on the size of the pre-test are the same as the effect of AR(1) errors, although the distortion observed is less.

3.4 Concluding Remarks

It is apparent that the presence of a non-scalar error covariance matrix is common in applied situations. In addition, if specification tests are carried out prior to testing or allowing for such a problem, the size of such preliminary tests may be distorted. The effect of this distortion and of the mis-specification itself on the estimators applied in such a situation is the topic we consider in the remainder of this thesis.

CHAPTER FOUR

PRE-TEST ESTIMATION OF THE COEFFICIENT VECTOR IN A LINEAR MODEL WITH A MIS-SPECIFIED ERROR COVARIANCE MATRIX

4.1 Introduction

In the following discussion we consider the properties of the classical pre-test estimator in the linear restriction model where, although the model is correctly specified in terms of the included regressors and error distribution, it is mis-specified with respect to the form of the error covariance matrix. The model to be estimated is

$$y = X\beta + \varepsilon ; \quad \varepsilon \sim N(0, \sigma^2 \Omega)$$

where y , X , β and ε are all defined as above and the matrix Ω is a general $T \times T$ positive definite symmetric matrix. The restrictions to be tested are represented by null hypotheses $H_0: R\beta = r$ vs $H_A: R\beta \neq r$ and the level of hypothesis error is defined as $\delta = R\beta - r$.

We will assume that the fitted model is

$$y = X\beta + \varepsilon ; \quad \varepsilon \sim N(0, \sigma^2 I_T),$$

the classical linear model, and that the usual unrestricted, OLS, and restricted, RLS, estimators are applied to it. The OLS and RLS estimators are defined as $b = S^{-1}X'y$ and $b^* = b + S^{-1}R'(RS^{-1}R')^{-1}(r - Rb)$ respectively, where $S = X'X$. The test statistic used to determine the validity of the restrictions is the usual Wald statistic

$$u = \frac{(Rb - R)(RS^{-1}R')^{-1}(Rb - r)(T - K)}{(y - Xb)'(y - Xb) - J},$$

and the null hypothesis is rejected if the test statistic is greater than the critical value, c , which is chosen such that

$$\Pr. \left\{ F_{(J, \nu)} < c \right\} = (1-\alpha),$$

where α is the nominal size of the pre-test and $\nu = T-K$. If the restrictions are rejected, the coefficient vector is estimated using the unrestricted estimator, b , otherwise the restricted estimator, b^* , is used. This procedure gives rise to the pre-test estimator $\hat{\beta}$ which is defined as

$$\hat{\beta} = \begin{cases} b & ; \text{ if } u \geq c \\ b^* & ; \text{ if } u < c \end{cases}$$

4.2 Properties of the Component Estimators and the Test Statistic

It is well known (see, for example, Judge *et al.* (1985)), that the OLS estimator is not biased by the mis-specification of the error covariance matrix and that its bias and risk, under quadratic loss, are given by

$$B(b, \beta) \equiv E(b) - \beta = 0$$

and

$$\begin{aligned} \rho(b, \beta) &\equiv E \left[\left(b - E(b) \right)' \left(b - E(b) \right) \right] \\ &= \sigma^2 \text{tr}(S^{-1} X' \Omega X S^{-1}) \end{aligned}$$

respectively.

Similarly, the bias and risk of the RLS estimator under quadratic loss are given by

$$\begin{aligned} B(b^*, \beta) &= -S^{-1} R' (R S^{-1} R')^{-1} (R \beta - r) \\ &= -\eta \delta \end{aligned}$$

and

$$\rho(b^*, \beta) = \rho(b) - \sigma^2 \text{tr}(2S^{-1} X' \Omega X C - C X' \Omega X C) \sigma^2 + \delta' \eta' \eta \delta, \text{ respectively,}$$

where $C = S^{-1} R' (R S^{-1} R')^{-1} R S^{-1}$.

As is the case where the model is well specified, the RLS estimator will generally be biased if the restrictions are not valid. In fact, the bias in the estimator is not affected by the mis-specification. Under the null hypothesis the RLS estimator will have a smaller risk than the OLS estimator if $\text{tr}(2S^{-1}X'\Omega XC - CX'\Omega XC)$ is positive. This always holds true if $\Omega = I_T$, however it does not necessarily hold for general Ω . In fact, the RLS estimator may have a greater risk than the OLS estimator when $\delta = 0$, depending on the data and the degree of mis-specification in the model. As the level of hypothesis error, δ , increases the risk of the RLS estimator increases without bound.

Because Ω is a symmetric matrix of full rank, there exists a symmetric invertible matrix $\Omega^{1/2}$ such that $\Omega^{1/2}\Omega^{1/2} = \Omega$. Now, consider the $T \times T$ symmetric matrix $\Phi \equiv \Omega^{1/2} \left(X C X' - \frac{J_C}{(T-K)} M \right) \Omega^{1/2}$, where $M = (I_T - X S^{-1} X')$. This matrix has T real eigenvalues, denoted λ_i , $i = 1, \dots, T$, and a $T \times T$ orthonormal eigenvector matrix Υ . The eigenvector corresponding to λ_i is the i 'th column of Υ , denoted v_i .

Theorem 4.1

Under the stated assumptions, the cumulative distribution function of the test statistic, u , is given by

$$\Pr.(u < c) = \Pr.(z' \Lambda z < 0)$$

$$= \Pr. \left(\sum_{i=1}^T \lambda_i \chi_{(1; \theta_i)}^2 < 0 \right),$$

where $\Lambda = \text{diag}(\lambda_i)$, $z = \sigma^{-1} \Upsilon' \Omega^{-1/2} (\epsilon + X \eta \delta)$ and the non-centrality parameters of each of the T , independent, χ^2 random variates are $\theta_i = \frac{1}{2\sigma^2} (v_i' \Omega^{-1/2} X \eta \delta)^2$ or $\theta_i = \frac{1}{2} \zeta_i^2$, where $\zeta_i = \sigma^{-1} v_i' \Omega^{-1/2} X \eta \delta$.

Proof.

See Appendix 4A.

Note that z is $N(\sigma^{-1}\Gamma'\Omega^{-1/2}X\eta\delta, I_T) = N(\zeta, I_T)$ distributed, where $\zeta = \sigma^{-1}\Gamma'\Omega^{-1/2}X\eta\delta$, therefore z_i^2 is independent of z_j^2 , $i \neq j$, where z_i and z_j denote the i 'th and j 'th elements of the $T \times 1$ vector z .

Using the above theorem, we can calculate the true size and power of this pre-test using the algorithm of Davies (1980).

As we would expect, the cumulative distribution function of the test statistic collapses to the c.d.f. of a $F'_{(J, \nu; \lambda)}$ distributed random variable when $\Omega = I_T$; i.e. when there is no mis-specification in the model.

To show this we require the following:

Lemma 4.1

The matrices $\sigma^2 \left(X'CX' - \frac{cJ}{(T-K)}M \right)$, $X'CX'$ and M have a common eigenvector matrix denoted \bar{Y} , with i 'th element \bar{y}_i . The eigenvalues of $\sigma^2 \left(X'CX' - \frac{cJ}{T-K}M \right)$ are denoted $\bar{\lambda}_i$ and may be ordered such that

$$\bar{\lambda}_i = \sigma^2 \quad \text{for } i = 1, \dots, J,$$

$$\bar{\lambda}_i = 0 \quad \text{for } i = J+1, \dots, K,$$

$$\bar{\lambda}_i = -\sigma^2 \frac{cJ}{T-K} \quad \text{for } i = K+1, \dots, T.$$

Proof

See Appendix 4A.

Lemma 4.2

When $\Omega = I_T$

$$\zeta_i = 0 \quad ; \quad \text{for } i = J+1, \dots, T,$$

therefore

$$\theta_1 = 0 \quad ; \quad \text{for } i = J+1, \dots, T.$$

Proof.

See Appendix 4A.

Corollary 4.1

When $\Omega = I_T$, the test statistic is distributed as an $F'_{(J, \nu; \lambda)}$ random variable.

Proof

See Appendix 4A.

In general, if $XCX'\Omega$ and $M\Omega$ are idempotent and orthogonal, then u is distributed as a central F random variable under the null hypothesis (see Searle (1982 p.356)). However, when $\Omega \neq I_T$, it is likely to be the case that $XCX'\Omega$ and $M\Omega$ are neither idempotent nor independent and u will not generally be distributed as a central F random variable.

In the well-specified pre-test problem, the test statistic is $F'_{(J, \nu; \lambda)}$ distributed, as noted above, and the overall level of hypothesis error in the model is represented by the non-centrality parameter λ . If the error covariance matrix is mis-specified, the test statistic is not $F'_{(J, \nu; \lambda)}$ distributed in general but is distributed as a weighted sum of $\chi^2_{(1; \theta_1)}$ random variables. We can define a scalar measure of the level of hypothesis error in the model as

$$\theta = \sum_{(1: \lambda_1 > 0)} \theta_1 = \sum_{(1: \lambda_1 > 0)} \frac{1}{2\sigma^2} (v_1 \Omega^{-1/2} X \eta \delta)^2.$$

If the restrictions are valid $\theta = 0$. However, as the level of hypothesis error, δ , increases θ increases without bound. Note that if $\Omega = I_T$, this expression for θ collapses to the expression for λ , the numerator non-centrality parameter.

4.3 The Bias and Risk of the Pre-Test Estimator

Theorem 4.2

Under the stated assumptions, the bias and risk functions of the PTE are

$$B(\hat{\beta}, \beta) = -S^{-1}R(RS^{-1}R')^{-1}RS^{-1}X'\Omega^{1/2}\Upsilon P_3 \Upsilon' \Omega^{-1/2}X\eta\delta$$

and

$$\begin{aligned} \rho(\hat{\beta}, \beta) &= \sigma^2 \text{tr}(S^{-1}X'\Omega XS^{-1}) - 2\sigma^2 \text{tr}(CX'\Omega^{1/2}\Upsilon B \Upsilon' \Omega^{1/2}XS^{-1}) \\ &\quad + 2\delta\eta' CX'\Omega^{1/2}\Upsilon P_3 \Upsilon' \Omega^{-1/2}X\eta\delta \\ &\quad + \sigma^2 \text{tr}(CX'\Omega^{1/2}\Upsilon B \Upsilon' \Omega^{1/2}XC) \end{aligned}$$

respectively, where B is a T × T matrix with ij'th element

$$B_{ij} = \begin{cases} P_{31} + \zeta_1^2 P_{51} & ; \text{ when } i = j \\ \zeta_1 \zeta_j P_{313j} & ; \text{ when } i \neq j \end{cases}$$

and

$$P_3 = \text{diag}\{P_{31}\},$$

with

$$P_{mi} = \Pr. \left(\lambda_i \chi_{(m; \theta_i)}^{2'} + \sum_{j \neq i}^T \lambda_j \chi_{(1; \theta_j)}^{2'} < 0 \right); \quad m = 3, 5$$

and

$$P_{313j} = \Pr. \left(\lambda_i \chi_{(3; \theta_i)}^{2'} + \lambda_j \chi_{(3; \theta_j)}^{2'} + \sum_{k \neq 1, j}^T \lambda_k \chi_{(1; \theta_k)}^{2'} < 0 \right); \quad i, j = 1, \dots, T$$

Proof.

See Appendix 4A.

The bias and risk of the PTE depend on the form of the data, the form of the restrictions, the true error covariance matrix and the critical value chosen for the test.

Corollary 4.2

As the critical value, c , tends towards 0 (∞) the bias and risk of the PTE tend towards the bias and risk of the unrestricted (restricted) estimator, as expected.

Proof.

See Appendix 4A.

In general, however, the bias and risk of the PTE differ from the biases and risks of its component estimators.

Lemma 4.3

When $\Omega = I_T$

- a $P_{31} = h(2,0) \quad ; \quad i = 1, \dots, J$
- b $P_{5i} = h(4,0) \quad ; \quad i = 1, \dots, J$
- c $P_{313j} = h(4,0) \quad ; \quad i = 1, \dots, J, i \neq j.$

where $h(i,j) = \Pr. \left(\frac{\chi^2_{(J+1;\lambda)}}{\chi^2_{(\nu+j)}} < \frac{cJ}{T-K} \right).$

$$= \Pr. \left(\chi^2_{(J+1;\lambda)} - \frac{cJ}{T-K} \chi^2_{(\nu+j)} < 0 \right).$$

Proof.

See Appendix 4A.

Using Lemmas 4.1 and 4.2 and 4.3 it is straightforward to show that the bias and risk of the PTE collapse to the corresponding functions for a correctly specified model, as given by Judge and Bock (1978, pp.90 and 101).

Corollary 4.3

When $\Omega = I_T$ the bias and risk of the PTE are given by¹

¹ This accords with Judge and Bock (1978, p.101 eq. 4.6.5).

$$B(\hat{\beta}, \beta) = -\eta \delta h(4,0)$$

and²

$$\rho(\hat{\beta}, \beta) = \sigma^2 \text{tr}(S^{-1}) - \sigma^2 \text{tr}(C) h(2,0) - \delta' \eta' \eta \delta \left[h(4,0) - h(2,0) \right] .$$

Given the complexity of the expressions for the pre-test estimator risk and bias, it is difficult to make further observations regarding them without numerical evaluation. This has been done using the SHAZAM package (White *et al.* (1990)) and a FORTRAN program written by the author. In both of these approaches, the Davies (1980) algorithm was used and both were executed on a VAX 6340 computer.

4.4 Numerical Evaluation

4.4.1 The Models

We consider the effects of the mis-specification of the error covariance matrix on the OLS, RLS and PTE risk functions in a number of regression models. As the risk formulae are data dependent, several different data sets are used in each model. The data are described in Appendix 4B.

Autoregressive errors

First, we consider the problem of testing the significance of one or more of the regressors in the quarterly regression model

$$y_t = x_t \beta + u_t \quad ; \quad t = 1, \dots, T.$$

where x_t is the t 'th row of the regressor matrix X and u_t is generated by either a simple AR(1) process, a simple AR(4) process or a simple MA(1)

² This is equivalent to the expression given in Judge and Bock (1978, p.92 eq. 4.3.13b) as in this case $W = I_K$ as we are considering unweighted risk.

process. These can be written

$$(1-\rho_1 L)u_t = \varepsilon_t$$

$$(1-\rho_4 L^4)u_t = \varepsilon_t$$

and

$$u_t = \rho \varepsilon_{t-1} + \varepsilon_t$$

respectively, where $-1 < \rho_1, \rho_4, \rho < 1$, $\varepsilon \sim N(0, \sigma^2 I_T)$ and L represents the lag operator such that $L^j u_t = u_{t-j}$. The restrictions are written in the usual way as $H_0: R\beta = r$ vs $H_A: R\beta \neq r$, and the test statistic used is the Wald statistic which, in the presence of the mis-specification, will be distributed as described in Section 3.2.

Heteroscedastic errors

Model a). We consider the application of the Chow test for structural change in the model.

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \equiv x\beta + \varepsilon$$

where y_1 and y_2 are $T_1 \times 1$ and $T_2 \times 1$ vectors of observations, X_1 and X_2 are $T_1 \times K/2$ and $T_2 \times K/2$ non-stochastic regressor matrices of full rank and $\varepsilon \sim N\left(0, \sigma^2 \begin{bmatrix} I_{T_1} & 0 \\ 0 & \psi I_{T_2} \end{bmatrix}\right)$, where $T_1 + T_2 = T$.

The parameter ψ measures the degree of heteroscedasticity in the errors. When $\psi = 1$ the errors are homoscedastic. In this case, the null and alternative hypotheses are $H_0: \beta_1 = \beta_2$ vs $H_A: \beta_1 \neq \beta_2$. This is the model considered by Toyoda (1974), Schmidt and Sickles (1977) and Giles and Lieberman (1991a). In this study we will consider the case where $T_1 = T_2 = \frac{1}{2}T$.

The other five heteroscedastic models consider the effects of testing exclusion restrictions in the linear model $y_t = x_t\beta + \varepsilon_t$, $t = 1, \dots, T$, where x_t is the t 'th row of the regressor matrix X which is $T \times K$, non-stochastic and of rank K . The error variance is assumed to be some function of a variable ω_t , not necessarily a regressor. The exclusion restrictions are written $H_0: R\beta = r$ vs $H_A: R\beta \neq r$ and are tested using the usual Wald statistic.

The five models differ in the functional form of the error covariance matrix. If we define α as a parameter of the functional form, tr as a linear trend variable and x as one of the columns of the X matrix; i.e., a regressor, the five functional forms are:

$$\text{Model b) } \text{covar}(\varepsilon) = \sigma^2 \begin{bmatrix} I_{T_1} & 0 \\ 0 & \psi I_{T_2} \end{bmatrix}; \quad T_1 = T_2 = \frac{1}{2}T.$$

$$\text{Model c) } \text{var}(\varepsilon_t) = f_c(z_t) = \sigma^2 x_t^\alpha,$$

$$\text{Model d) } \text{var}(\varepsilon_t) = f_d(z_t) = \sigma^2 \exp(\alpha x_t)$$

$$\text{Model e) } \text{var}(\varepsilon_t) = f_e(z_t) = \sigma^2 tr_t^\alpha$$

$$\text{Model f) } \text{var}(\varepsilon_t) = f_f(z_t) = \sigma^2 \exp(\alpha tr_t).$$

For models (c) to (f) the measure of heteroscedasticity, ψ , is defined as being

$$\psi = \frac{f_i \left(\max_{t=1, \dots, T} (z_t) \right)}{f_i \left(\min_{t=1, \dots, T} (z_t) \right)} ; \quad i = c, d, e \text{ or } f.$$

A value of ψ of greater than unity implies that the error variance is increasing as z_t increases, while a ψ value of less than unity implies that the error variance is decreasing as z_t increases.

4.4.2 Numerical Results

The nominal test size is fixed at 5% in the discussion which follows. There is no size correction applied to the test as we wish to determine the consequences of assuming that the error term is well behaved when in fact it may not be. For the purposes of this discussion the term "power" refers to the size uncorrected power of the test.

Typical OLS, RLS and PTE risk functions for a regression model that is correctly specified (i.e., $\psi = 1$, $\rho_1, \rho_4, \rho = 0$) are shown in Figures 4.1, 4.3 and 4.5. Quantitatively, the presence of an autoregressive, moving average or heteroscedastic process in the error term may increase, or slightly reduce, the risks of the estimators for each value of θ . Qualitatively, the mis-specification introduces a bias in the pre-test power function and changes the relative dominance of the three estimators.

The effect of the mis-specification on the true size and power of the pre-test depends on a number of different factors, including the number of regressors in the model, the particular characteristics of the regressors and the form of the true error covariance matrix. For example, if the significance of a (group of) trended regressors is being tested, the true size and power of the test increase with increasing ρ , in the case of MA(1) errors, or with increasing ρ_1 , in the case of AR(1) errors.³ The converse may occur if the regressors are not trended. As the value of ρ_1 or ρ , decreases below zero the opposite effect is observed with the (true) size and

³ A similar result is found by Consiglieri (1981) and Giles and Scott (1992) in the case of the size of the Chow test, and this is consistent with Kiviet's (1980) results.

power of the test falling for most regressor sets.⁴

In general, if the errors are generated by an AR(4) process, the power of the pre-test is reduced if the absolute value of ρ_4 is close to unity. An exception to this is the case of testing the joint significance of a set of seasonal dummy variables where the true size and power of the test increase with increasing values of ρ_4 . The opposite effect is observed as ρ_4 decreases below zero. Figure 4.2 illustrates the effects of a downward distortion in the power function on the PTE risk. Comparing Figure 4.2 with Figure 4.1 we see that the PTE risk is closer to the RLS risk at each level of θ than is assumed to be the case.

When the errors are heteroscedastic, an increase in the degree of heteroscedasticity appears likely to increase the true size and power of the test for small values of θ , and to reduce the power for large values of θ , if there are three or more regressors. In models with less than three regressors, there is no consistent pattern.

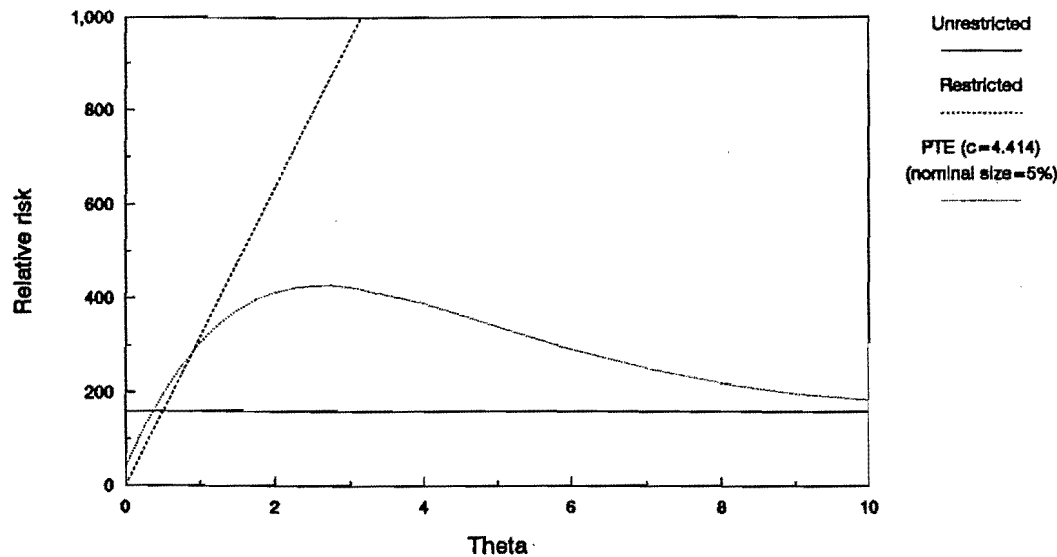
When the errors are generated by an AR(1) process, any increase in the value of ρ_1 generally has the effect of decreasing the range of θ over which it is preferable to pre-test rather than to simply ignore prior information and estimate using OLS. With some regressor sets OLS may strictly dominate both the RLS and the PT estimators. In this case the imposition of valid restrictions serves to increase the estimator risk.⁵

⁴ This is also consistent with Giles and Scott's (1992) results.

⁵ A similar result is obtained by Giles and Giles (1991) in the case of estimating the regression scale parameter if a sufficiently asymmetric loss function is used. It is also partially analogous to Mittelhammer's (1984) result that in a model mis-specified by the exclusion of relevant regressors the RLS and PT estimators may be dominated by OLS if the restrictions are valid.

FIGURE 4.1: Coefficient Estimator Risk Functions

Well Specified Model
Regressors, constant and Australian C.P.I.

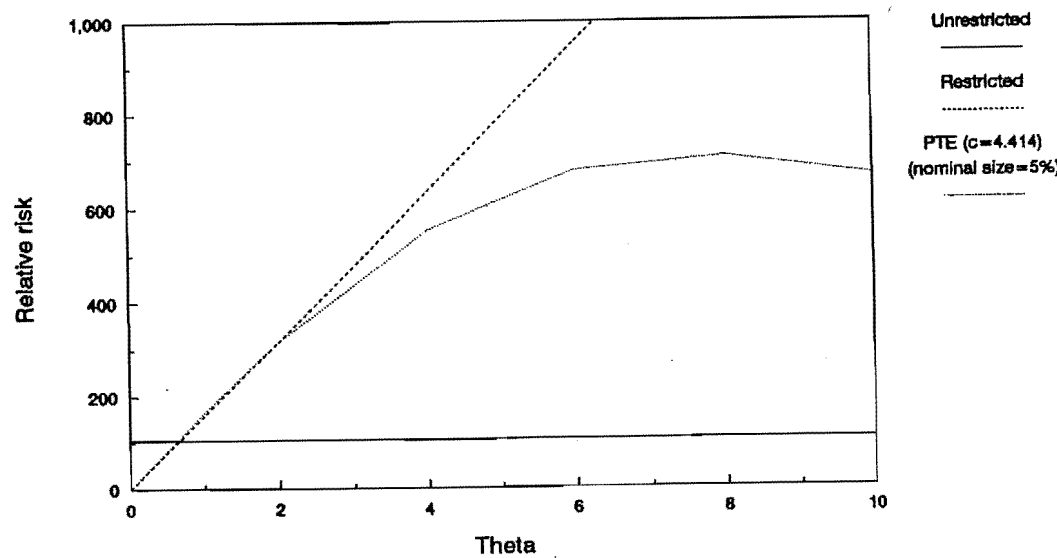


One restriction

Sample Size = 20

FIGURE 4.2: Coefficient Estimator Risk Functions

MA(1) errors; $\rho = -0.9$
Regressors, constant and Australian C.P.I.



One restriction

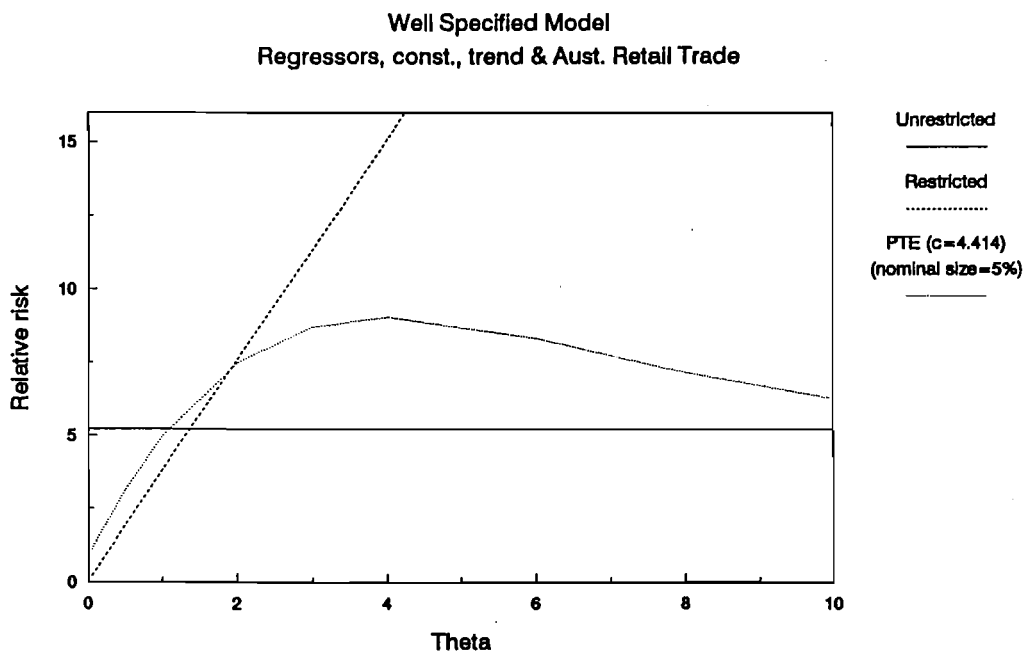
Sample Size = 20

Conversely, if ρ_1 decreases the range of θ over which it is preferable to pre-test generally increases, making the use of the PTE more attractive relative to OLS. An example of this is shown in Figure 4.4. Comparing Figure 4.4 with Figure 4.3 we see that the PTE dominates OLS over a larger part of the θ space than is assumed to be the case and also that there exists a range of θ over which the PTE dominates both of its component estimators. This is in contrast to the usual result, that in a correctly specified model the PTE of β is never the minimum risk estimator and will have higher risk than both of its two component estimators over some part of the θ range.⁶ When the errors are generated according to a moving average process, increasing values of ρ appear to have little or no effect on the relative dominance of the estimators.

In the models with heteroscedastic errors, increasing levels of heteroscedasticity are more likely to increase the range of θ over which the PTE risk is lower than the OLS risk than to reduce it, particularly if the regressor variables are trended. There are instances, however, when the converse occurs. In some cases the RLS and PT estimators may become completely dominated by OLS as shown in Figure 4.6. However, no general result is apparent as the effect of the mis-specification varies with both the type of heteroscedasticity and the form of the regressor variables.

⁶ Another example of the possible dominance of a PTE is given in Ohtani (1983), who shows that, in a model in which a relevant, unobservable, variable has been replaced by a proxy variable, there may exist a region in which the PTE risk is lower than the risks of its two component estimators. Other examples are given in Giles (1991a & b), Giles and Giles (1991), and Wong and Giles (1991).

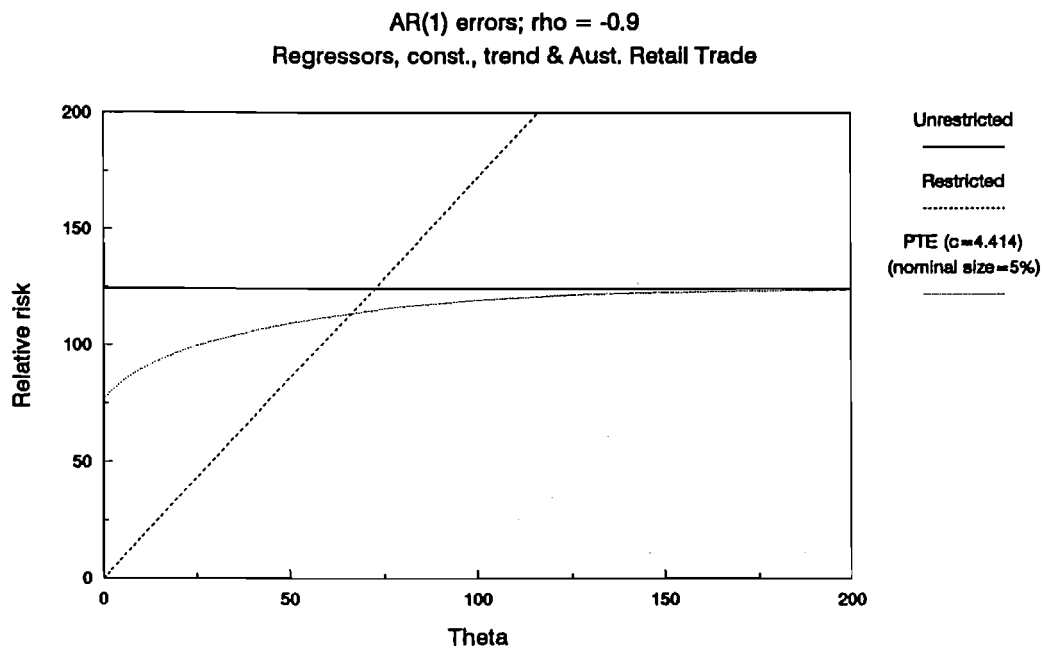
FIGURE 4.3: Coefficient Estimator Risk Functions



Joint test of linear trend and Retail Trade

Sample Size = 20

FIGURE 4.4: Coefficient Estimator Risk Functions

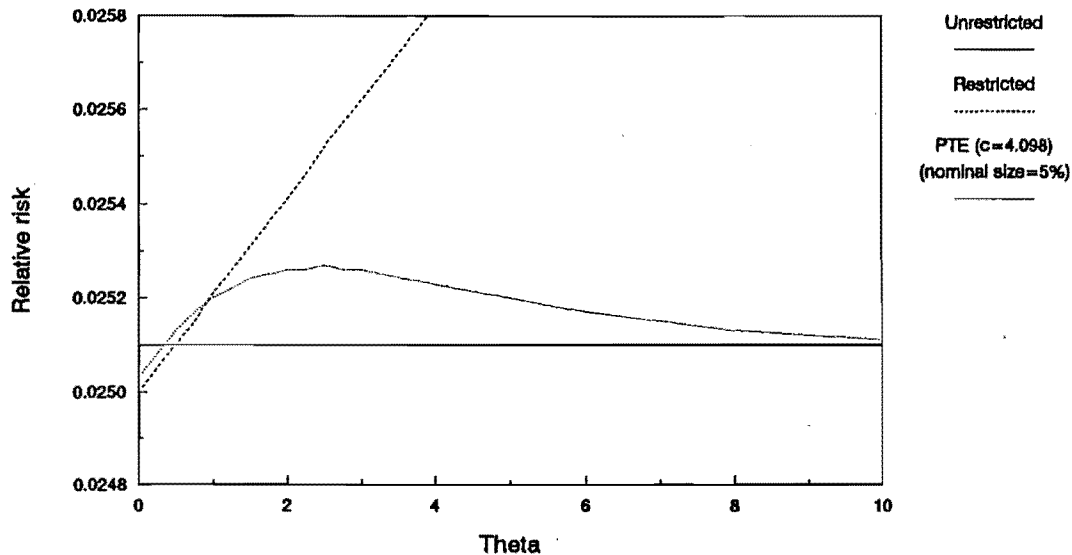


Joint test of linear trend and Retail Trade

Sample Size = 20

FIGURE 4.5: Coefficient Estimator Risk Functions

Well Specified Model
Regressors, constant and $N(0,1)$ r.v.

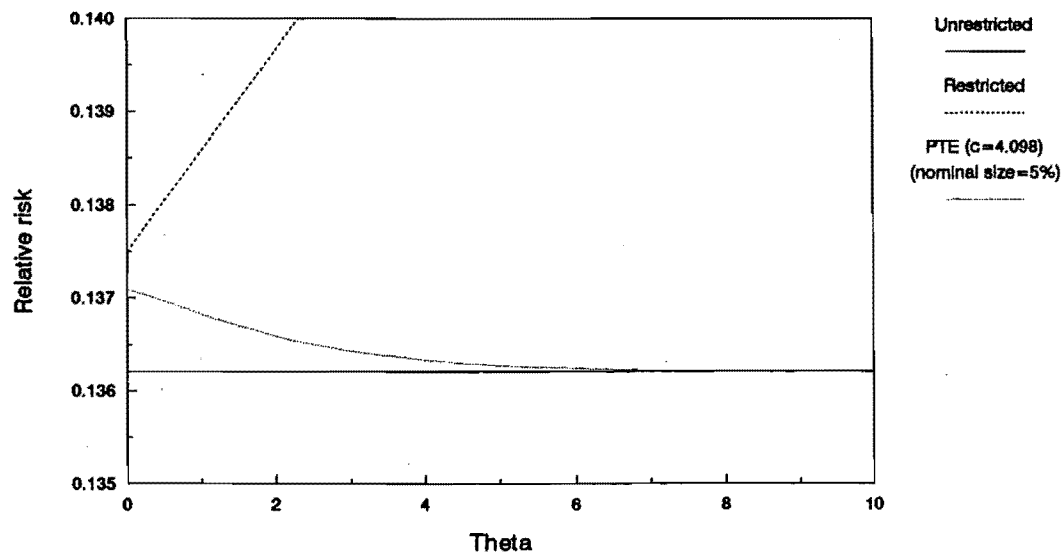


One restriction

Sample Size = 40

FIGURE 4.6: Coefficient Estimator Risk Functions

Heteroscedastic Errors (Model B); $\psi = 10.0$
Regressors, constant and $N(0,1)$ r.v.



One restriction

Sample Size = 40

Other things being equal, an increase in the sample size leads to a reduction in estimator risk for each value of θ . However such an increase may not alleviate the distortions introduced to the models as a result of the mis-specification. The (size uncorrected) power of the test may, in fact, be reduced by such an increase if the mis-specification is severe and θ is close to zero. Also, the range of θ over which the PTE has a lower risk than the OLS estimator may be further reduced or increased by an increase in sample size.

In general the effects of increasing the sample size are ambiguous, particularly with real or non-trended data. This may be because the additional data points may change the characteristics of the regressor set and all of the models considered are very sensitive to the form of the data.

4.5 Conclusion

The practical implications of the broad type of model mis-specification that we are considering vary depending on the specific type of mis-specification and the regressors. Because of this, little can be offered by way of a general prescription. Some points, however, can be made.

When the errors are generated by an AR or MA process with positive coefficients, and the regressors whose coefficients are included in the restrictions are trended, the PTE may be strictly dominated by OLS, in which case it is better to ignore the prior information. Even if the PTE is not strictly dominated, the θ range over which the risk of the PTE is lower than OLS is generally reduced (compared to the correctly specified model). The regret associated with using OLS rather than the PTE in that θ range is reduced also, as the pre-test power function is likely to be distorted upwards and the test will tend to over-reject valid restrictions. Hence,

although in practice the degree of distortion is unknown, it may be preferable to ignore the prior information rather than pre-test if an autocorrelation problem is suspected.

Conversely, if the errors are generated by an AR or MA process with negative coefficients, or if the regressors whose coefficients are included in the restrictions are not trended, the power of the pre-test is likely to be reduced by the mis-specification and the PTE will dominate OLS over a greater portion of the θ range compared to the correctly specified model. Therefore it is likely to be preferable to pre-test rather than ignore the prior information in this case.

On the basis of these results, it appears to be advisable to test for such processes before any testing of the validity of linear restrictions is carried out. If the linear restrictions involve the coefficients of trended regressors, it may be wise to choose a critical value such that the test has a higher power against a positive AR or MA process than against a negative process, as the costs of failing to correct for a positive process are the greater of the two. The converse is true if the restrictions involve the coefficients of non-trended regressors. However, it should be noted that there are further implications associated with such multiple pre-testing (e.g., see King and Giles (1984) and Giles and Lieberman (1992a)).

If the errors are possibly heteroscedastic, there is no general prescription as, although increasing heteroscedasticity may increase the pre-test size, the power of the pre-test will be reduced in models with more than two regressors. Also, as we have seen, the PTE may be strictly dominated by OLS, in which case the prior information should be ignored. However, given that the true error covariance matrix is unknown, the effect of the mis-specification on a given model cannot be determined.

Because of the effects of the mis-specification on the (true) size and power of the test, any attempt to apply an "optimal" critical value, such as is suggested by Brook (1976), will not necessarily lead to an "optimal" pre-test risk. We examine this issue further in Chapter Eight.

Appendix 4A

Proof of Theorem 4.1.

Define the normal random vector $\bar{\varepsilon} \equiv (\varepsilon + X\eta\delta)$, $\bar{\varepsilon} \sim N(X\eta\delta, \sigma^2\Omega)$, then we have

$$u = \bar{\varepsilon}' X C X' \bar{\varepsilon} (T-K) / \bar{\varepsilon}' M \bar{\varepsilon} J$$

as

$$\begin{aligned} \bar{\varepsilon}' M \bar{\varepsilon} &= (\delta' \eta' X' + \varepsilon') M (\varepsilon + X\eta\delta) \\ &= \varepsilon' M \varepsilon \\ &= (y - Xb)' (y - Xb) \end{aligned}$$

and

$$\begin{aligned} \bar{\varepsilon}' X C X' \bar{\varepsilon} &= (\delta' \eta' X' + \varepsilon') X C X' (\varepsilon + X\eta\delta) \\ &= \varepsilon' X C X' \varepsilon + 2\varepsilon' X S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X' X S^{-1} R' (R S^{-1} R')^{-1} \delta \\ &\quad + \delta' (R S^{-1} R')^{-1} R S^{-1} X' X S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X' X S^{-1} R' (R S^{-1} R')^{-1} \delta \\ &= \varepsilon' X S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} \varepsilon + 2\varepsilon' X S^{-1} R' (R S^{-1} R')^{-1} \delta \\ &\quad + \delta' (R S^{-1} R')^{-1} \delta \\ &= (\delta + R S^{-1} X' \varepsilon)' (R S^{-1} R')^{-1} (\delta + R S^{-1} X' \varepsilon) \\ &= \left(R S^{-1} X' (X\beta + \varepsilon) - r \right)' (R S^{-1} R')^{-1} \left(R S^{-1} X' (X\beta + \varepsilon) - r \right) \\ &= (Rb - r)' (R S^{-1} R')^{-1} (Rb - r). \end{aligned}$$

Hence, following the method of Koerts and Abrahamse (1969),

$$\begin{aligned} \Pr.(u < c) &= \Pr. \left(\bar{\varepsilon}' X C X' \bar{\varepsilon} / \bar{\varepsilon}' M \bar{\varepsilon} < \frac{cJ}{T-K} \right) \\ &= \Pr. \left(\bar{\varepsilon}' X C X' \bar{\varepsilon} - \frac{cJ}{T-K} \bar{\varepsilon}' M \bar{\varepsilon} < 0 \right) \\ &= \Pr. \left(\bar{\varepsilon}' \Omega^{-1/2} \Upsilon \Upsilon' \Omega^{1/2} (X C X' - \frac{cJ}{T-K} M) \Omega^{1/2} \Upsilon \Upsilon' \Omega^{-1/2} \bar{\varepsilon} < 0 \right) \\ &= \Pr.(z' \Lambda z < 0) \text{ by definition.} \end{aligned}$$

□

Proof of Lemma 4.1.

The $T \times T$ symmetric, positive semi-definite matrix XCX' has rank of J and, because it is an idempotent matrix, it has J unit eigenvalues and corresponding eigenvectors v_1^C , $i = 1, \dots, J$. Therefore

$$XCX' v_1^C = v_1^C \quad ; \quad i = 1, \dots, J \text{ by definition}$$

hence

$$XCX' v_1^C = XS^{-1}X' v_1^C \quad ; \quad i = 1, \dots, J$$

or

$$XCX' v_1^C = XS^{-1}X' v_1^C = v_1^C \quad ; \quad i = 1, \dots, J.$$

Therefore

$$Mv_1^C = I_T v_1^C - XS^{-1}X' v_1^C = 0 \quad ; \quad i = 1, \dots, J.$$

Therefore each v_1^C is a unit eigenvector of $\left(XCX' - \frac{CJ}{T-K}M\right)$ as it is a unit eigenvector of XCX' and a zero eigenvector of M . Let $\bar{v}_1^C = v_1^C$; $i = 1, \dots, J$.

Now consider the $T \times T$ symmetric, idempotent, positive semi-definite matrix $A \equiv (XS^{-1}X' - XCX')$. This matrix is of rank $K-J$ and therefore has $K-J$ unit eigenvectors, denoted v_1^A , $i = 1, \dots, K-J$.

$$Av_1^A = v_1^A \quad ; \quad i = 1, \dots, K-J \text{ by definition}$$

or

$$XS^{-1}X' v_1^A - XCX' v_1^A = v_1^A \quad ; \quad i = 1, \dots, K-J,$$

therefore

$$XS^{-1}X' XS^{-1}X' v_1^A - XS^{-1}X' XCX' v_1^A = XS^{-1}X' v_1^A \quad ; \quad i = 1, \dots, K-J.$$

or

$$\Rightarrow Av_1^A = XS^{-1}X' v_1^A = v_1^A \quad ; \quad i = 1, \dots, K-J.$$

Therefore, each v_1^A , $i = 1, \dots, K-J$ is a unit eigenvector of $XS^{-1}X'$ and a zero eigenvector of XCX' , M and $\left(XCX' - \frac{cJ}{(T-K)}M\right)$. Let $\bar{v}_{(J+1)} = v_1^A$; $i = 1, \dots, K-J$.

Finally, consider the $T \times T$, symmetric, idempotent, positive semi-definite matrix M . This matrix is of rank $T-K$ and has $T-K$ unit eigenvectors, denoted v_1^M ; $i = 1, \dots, T-K$.

$$(I_T - XS^{-1}X')v_1^M = v_1^M \quad ; \quad i = 1, \dots, T-K \text{ by definition.}$$

Therefore

$$X'(I_T - XS^{-1}X')v_1^M = 0v_1^M = X'v_1^M \quad ; \quad i = 1, \dots, T-K.$$

Therefore, each v_1^M is a zero eigenvector of XCX' and an eigenvector of $\left(XCX' - \frac{cJ}{T-K}M\right)$ corresponding to a root of $-\frac{cJ}{T-K}$. Let $\bar{v}_{(K+1)} = v_1^M$; $i = 1, \dots, T-K$.

The matrix \bar{Y} , the first J rows of which comprise the unit eigenvectors of XCX' , the next $K-J$ rows comprise the unit eigenvectors of $(XS^{-1}X' - XCX')$ and the last $T-K$ rows comprise the unit eigenvectors of M , is a joint eigenvector matrix of $\sigma^2\left(XCX' - \frac{cJ}{T-K}M\right)$, XCX' and M by definition. \square

Proof of Lemma 4.2.

When $\Omega = I_T$ the eigenvectors of Φ are also eigenvectors of XCX' and in particular

$$XCX'v_1 = 0 \quad ; \quad \text{for } i = J+1, \dots, T,$$

by Lemma 4.1.

Now

$$\begin{aligned} \zeta_1 &= \sigma^{-1}\delta'(RS^{-1}R')^{-1}RS^{-1}X'v_1 \quad \text{when } \Omega = I_T \\ &= \sigma^{-1}\delta'(RS^{-1}R'S^{-1}RS^{-1}X'XS^{-1}R'(RS^{-1}R')^{-1}RS^{-1}X'v_1) \\ &= \sigma^{-1}\delta'\eta'X'XCX'v_1 = 0 \quad ; \quad \text{for } i = J+1, \dots, T. \end{aligned} \quad \square$$

Proof of Corollary 4.1.

Therefore, when $\Omega = I_T$,

$$\begin{aligned} & \Pr. \left(\sum_{i=1}^T \lambda_i \chi_{(1;\theta_i)}^{2'} < 0 \right) \\ &= \Pr. \left(\sum_{i=1}^J \sigma^2 \chi_{(1;\theta_i)}^{2'} + \sum_{i=J+1}^K 0 \chi_{(1;\theta_i)}^{2'} + \sum_{i=K+1}^T -\frac{cJ}{T-K} \sigma^2 \chi_{(1;\theta_i)}^{2'} < 0 \right), \end{aligned}$$

by Lemma 4.1. Now, $\theta_i = 0$, for $i = J+1, \dots, T$, by Lemma 4.2.

$$\begin{aligned} \text{Hence } \sum_{i=1}^J \theta_i &= \sum_{i=1}^T \theta_i = \frac{1}{2} \zeta' \zeta \\ &= \frac{1}{2\sigma^2} \left(\delta' (RS^{-1}R')^{-1} RS^{-1} X' \bar{Y} \bar{Y}' X S^{-1} R' (RS^{-1}R')^{-1} \delta \right) \\ &= \frac{1}{2\sigma^2} \delta' (RS^{-1}R')^{-1} \delta. \end{aligned}$$

Hence

$$\begin{aligned} & \Pr. \left(\sum_{i=1}^T \lambda_i \chi_{(1;\theta_i)}^{2'} < 0 \right) \\ &= \Pr. \left(\chi_{(J;\lambda)}^{2'} - \frac{cJ}{T-K} \chi_{(T-K)}^2 < 0 \right) \\ &= \Pr. \left(\frac{\chi_{(J;\lambda)}^{2'} (T-K)}{\chi_{(T-K)}^2 J} < c \right). \end{aligned}$$

□

Proof of Theorem 4.2.

To prove this theorem we require the following lemma.

Lemma 4A.1

Let Ψ be an indicator function such that $\Psi(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ 1 & \text{if } x < 0 \end{cases}$ then

- a) $E[x' \Psi(z' \Lambda z)] = P_3 \zeta$, and
- b) $E[zz' \Psi(z' \Lambda z)] = B$.

Proof

a) Consider the i 'th element of the $T \times 1$ vector $E[z\Psi(z'\Lambda z)]$,

$$E[z_i\Psi(z'\Lambda z)] = E\left[z_i\Psi\left(\lambda_1 z_1^2 + \sum_{j \neq 1}^T \lambda_j z_j^2\right)\right].$$

$\text{Covar}(z) = I_T$ and therefore z_i is independent of z_j , $i \neq j$. Hence

$$\begin{aligned} E[z_i\Psi(z'\Lambda z)] &= E_{j \neq 1} \left[E_i \left[z_i \Psi\left(\lambda_1 z_1^2 + \sum_{j \neq 1}^T \lambda_j z_j^2\right) \right] \right] \\ &= E_{j \neq 1} \left[E_i \left[\zeta_i \left(\lambda_1 \chi_{(3, \theta_1)}^2 + \sum_{j \neq 1}^T \lambda_j z_j^2 \right) \right] \right], \end{aligned}$$

by Lemma 2 of Judge and Bock (1978, p.320).

Now, because $E[\Psi(x)] = \Pr(x < 0)$ by definition,

$$E[z_i\Psi(z'\Lambda z)] = P_{3i} \zeta_i \quad i = 1, \dots, T,$$

therefore

$$E[z\Psi(z'\Lambda z)] = P_3 \zeta.$$

b) Consider the ij 'th element of the $T \times T$ matrix $E[zz'\Psi(z'\Lambda z)]$.

It is straightforward to show that when $i = j$,

$$E[z_i z_j \Psi(z'\Lambda z)] = P_{3i} + (\zeta_i)^2 P_{5i},$$

by Lemma 1 of Judge and Bock (1978, p.320), and that, when $i \neq j$,

$$E[z_i z_j \Psi(z'\Lambda z)] = (\zeta_i)(\zeta_j) P_{3i3j},$$

by Lemma 2 of Judge and Bock (1978, p.320). □

Proof of Theorem 4.2

$$\text{The PTE can be written } \hat{\beta} = \begin{cases} b & \text{if } u \geq c \\ b + S^{-1}R'(RS^{-1}R')^{-1}(r-Rb) & \text{if } u < c \end{cases}$$

Because $u < c \Leftrightarrow z'\Lambda z < 0$, we can write this as

$$\begin{aligned}
\hat{\beta} &= b + S^{-1}R'(RS^{-1}R')^{-1}(r-Rb)\Psi(z'\Lambda z). \\
&= \beta + S^{-1}X'\epsilon + \eta(r-R\beta-RS^{-1}X'\epsilon)\Psi(z'\Lambda z) \\
&= \beta + S^{-1}X'\epsilon - \sigma CX'\Omega^{1/2}\Upsilon z\Psi(z'\Lambda z).
\end{aligned}$$

The bias of the PTE is

$$\begin{aligned}
B(\hat{\beta}, \beta) &= E(\hat{\beta}) - \beta = S^{-1}X' E(\epsilon) - CX'\Omega^{1/2}\Upsilon E[z\Psi(z'\Lambda z)] \\
&= -S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}X'\Omega^{1/2}\Upsilon P_3\Upsilon'\Omega^{(-1/2)}X\eta\delta,
\end{aligned}$$

by Lemma 4A.1.

The risk of the PTE is

$$\begin{aligned}
\rho(\hat{\beta}, \beta) &= E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] \\
&= E\left[\epsilon'XS^{-1}S^{-1}X'\epsilon - 2\sigma\epsilon'XS^{-1}CX'\Omega^{1/2}\Upsilon z\Psi(z'\Lambda z) \right. \\
&\quad \left. + \sigma^2z'\Upsilon'\Omega^{(1/2)}XC\Psi(z'\Lambda z)CX'\Omega^{1/2}\Upsilon z\right].
\end{aligned}$$

Now, because $\epsilon = \sigma\Omega^{1/2}\Upsilon z - X\eta\delta$,

$$\begin{aligned}
\rho(\hat{\beta}, \beta) &= E\left[\sigma^2\text{tr}(\epsilon'XS^{-1}X'\epsilon) \right. \\
&\quad \left. - 2\sigma^2\text{tr}\left(z'\Psi'\Omega^{1/2}XS^{-1}CX'\Omega^{1/2}\Upsilon z\Psi(z'\Lambda z)\right) \right. \\
&\quad \left. + 2\sigma\delta'\eta'X'XS^{-1}CX'\Omega^{1/2}\Upsilon z\Psi(z'\Lambda z) \right. \\
&\quad \left. + \sigma^2\text{tr}\left(z'\Upsilon'\Omega^{1/2}XC\Psi(z'\Lambda z)\right)CX'\Omega^{1/2}\Upsilon z\right] \\
\rho(\hat{\beta}, \beta) &= \sigma^2\text{tr}\left(S^{-1}X'\Omega XS^{-1}\right) - 2\sigma^2\text{tr}\left(CX'\Omega^{1/2}\Upsilon B\Upsilon'\Omega^{1/2}XS^{-1}\right) \\
&\quad + 2\sigma\delta'\eta'CX'\Omega^{1/2}\Upsilon P_3\Upsilon'\Omega^{-1/2}X\eta\delta \\
&\quad + \sigma^2\text{tr}\left(CX'\Omega^{1/2}\Upsilon B\Upsilon'\Omega^{1/2}XC\right)
\end{aligned}$$

by Lemma 4A.1. □

Proof of Corollary 4.2

Consider the eigenvectors of $\Phi = \sigma^2 \Omega^{1/2} \left(X C X' - \frac{c J}{T-K} M \right) \Omega^{1/2}$. Because $\Omega^{1/2} X C X' \Omega^{1/2}$ and $\Omega^{1/2} M \Omega^{1/2}$ are positive semi-definite matrices they have only non-negative eigenvalues, therefore $\frac{c J}{(T-K)} \Omega^{1/2} M \Omega^{1/2}$ has only non-positive eigenvalues. Hence, as c tends towards zero, (ω) , the negative eigenvalues of Φ also tend towards zero (ω) . This has the effect of making the P_{31} , P_{51} , P_{31j} values tend towards zero (unity). In the former case it is obvious that the functions collapse to the corresponding functions for the unrestricted estimator. In the latter case we have

$$\lim_{c \rightarrow \infty} (B) = I_T + \zeta \zeta'$$

and

$$\lim_{c \rightarrow \infty} P_3 = I_T \quad \text{as } c \rightarrow \infty.$$

Therefore,

$$\begin{aligned} \lim_{c \rightarrow \infty} [B(\hat{\beta}, \beta)] &= -S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X' \Omega^{1/2} \Upsilon \Upsilon' \Omega^{-1/2} X \eta \delta \\ &= -S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X' X S^{-1} R' (R S^{-1} R')^{-1} \delta \\ &= -S^{-1} R' (R S^{-1} R')^{-1} \delta = B(b^*, \beta). \end{aligned}$$

Similarly,

$$\begin{aligned} \lim_{c \rightarrow \infty} [\rho(\hat{\beta}, \beta)] &= \sigma^2 \text{tr}(S^{-1} X' \Omega X S^{-1}) - 2\sigma^2 \text{tr} \left(C X' \Omega^{1/2} \Upsilon (I_T + \zeta \zeta') \Upsilon' \Omega^{1/2} X S^{-1} \right) \\ &\quad + 2\delta' \eta' C X' \Omega^{1/2} \Upsilon \Upsilon' \Omega^{-1/2} X \eta \delta \\ &\quad + \sigma^2 \text{tr} \left(C X' \Omega^{1/2} \Upsilon (I_T + \zeta \zeta') \Upsilon' \Omega^{1/2} X S^{-1} \right). \end{aligned}$$

Now,

$$\begin{aligned} \sigma^2 \text{tr} \left(C X' \Omega^{1/2} \Upsilon (I_T + \zeta \zeta') \Upsilon' \Omega^{1/2} X S^{-1} \right) &= \sigma^2 \text{tr} \left(C X' \Omega^{1/2} X S^{-1} \right) \\ &\quad + \delta' \eta' X' \Omega^{-1/2} \Upsilon \Upsilon' \Omega^{1/2} X S^{-1} C X' \Omega^{1/2} \Upsilon \Upsilon' \Omega^{-1/2} X \eta \delta. \\ &= \sigma^2 \text{tr} \left(C X' \Omega X S^{-1} \right) + \delta' \eta' \eta \delta. \end{aligned}$$

Similarly

$$\sigma^2 \text{tr} \left(CX' \Omega^{1/2} \Upsilon (I_T + \zeta \zeta') \Upsilon' \Omega^{1/2} XC \right) = \sigma^2 \text{tr} \left(CX' \Omega XC \right) + \delta' \eta' \eta \delta,$$

therefore

$$\begin{aligned} \lim_{c \rightarrow \infty} [\rho(\hat{\beta}, \beta)] &= \sigma^2 \text{tr} (S^{-1} X' \Omega X S^{-1}) - 2\sigma^2 \text{tr} \left(CX' X S^{-1} \right) - 2\delta' \eta' \eta \delta \\ &\quad + 2\delta' \eta' \eta \sigma + 2\sigma^2 \text{tr} \left(CX' \Omega XC \right) + \delta' \eta' \eta \delta \\ &= \rho(b^*, \beta). \end{aligned}$$

□

Proof of Lemma 4.3.

Recall the definition of P_3 :

$$\begin{aligned} P_{3i} &= \Pr. \left(\lambda_1 \chi_{(3; \theta_1)}^{2'} + \sum_{j \neq 1} \lambda_j \chi_{(1; \theta_j)}^{2'} < 0 \right), \\ &= \Pr. \left(\sigma^2 \chi_{(3; \theta_1)}^{2'} + \sum_{j \neq 1}^J \sigma^2 \chi_{(1; \theta_j)}^{2'} + \sum_{j=J+1}^K 0 \chi_1^2 - \sum_{j=K+1}^T \sigma^2 \frac{cJ}{T-K} \chi_1^2 < 0 \right), \end{aligned}$$

when $\Omega = I_T$ for $i = 1, \dots, J$ by Lemma 4.1 and 4.2.

Therefore $P_{3i} = \Pr. \left(\chi_{(J+2; \lambda)}^{2'} < \frac{cJ}{T-K} \chi_{(T-K)}^2 \right) = h(2, 0)$, for $i = 1, \dots, J$ when $\Omega = I_T$. Similarly $P_{5i} = h(4, 0)$ and $P_{3i3j} = (4, 0)$, for $i, j = 1, \dots, T$ when $\Omega = I_T$. □

Proof of Corollary 4.3.

Consider $P_3 \zeta_1$. By Lemma 4.2, $\zeta_1 = 0$ when $\Omega = I_T$ for $i \geq J + 1$, hence $P_3 \zeta_1 = h(2, 0) \zeta_1$. Therefore

$$\begin{aligned} B(\hat{\beta}, \beta) &= S^{-1} R' (RS^{-1} R')^{-1} RS^{-1} X' \Omega^{1/2} \Upsilon P_3 \Upsilon' \Omega^{-1/2} X \eta \delta \\ &= S^{-1} R' (RS^{-1} R')^{-1} RS^{-1} X' X S^{-1} R' (RS^{-1} R')^{-1} \delta h(4, 0) \\ &= -\eta \delta h(4, 0) \quad \text{when } \Omega = I_T. \end{aligned}$$

Now consider the matrix B . When $\Omega = I_T$, $B = \bar{B} + \zeta \zeta' h(4, 0)$, where \bar{B} is a diagonal matrix with elements $\bar{B}_{11} = P_{31}$. Recall that when $\Omega = I_T$, $\Upsilon = \bar{\Upsilon}$ an

orthonormal eigenvector matrix of XCX' with $\Upsilon'XCX'\Upsilon = \bar{I}_J \equiv \begin{bmatrix} I_J & 0 \\ 0 & 0 \end{bmatrix}$.

Therefore we have

$$\begin{aligned} CX'\Omega^{1/2}\Upsilon B &= S^{-1}X'\Upsilon\Upsilon'XCX'\Upsilon B \\ &= S^{-1}X'\Upsilon\bar{I}_J B \\ &= S^{-1}X'\Upsilon\left(\bar{I}_J h(2,0) + \zeta\zeta'h(4,0)\right), \quad \text{when } \Omega = I_T. \end{aligned}$$

Consider each of the terms in

$$\begin{aligned} \rho(\hat{\beta}, \beta) &= \sigma^2 \text{tr}(S^{-1}X'\Omega XS^{-1}) - 2\sigma^2 \text{tr}(CX'\Omega^{1/2}\Upsilon B \Upsilon' \Omega^{1/2}XS^{-1}) \\ &\quad + 2\delta'\eta'CX\Omega^{1/2}\Upsilon P_3\zeta \\ &\quad + \sigma^2 \text{tr}(CX'\Omega^{1/2}\Upsilon B \Upsilon' \Omega^{1/2}CX). \end{aligned}$$

When $\Omega = I_T$, $\text{tr}(S^{-1}X'\Omega XS^{-1}) = \text{tr}(S^{-1})$. Also, $\text{tr}(CX'\Omega^{1/2}\Upsilon B \Upsilon' \Omega^{1/2}XS^{-1}) = \text{tr}\left(S^{-1}X'\Upsilon\left(\bar{I}_J h(2,0) + \zeta\zeta'h(4,0)\right)\Upsilon'XS^{-1}\right) = \text{tr}(CX'\Omega^{1/2}\Upsilon B \Upsilon' \Omega^{1/2}XC)$.

Now, when $\Omega = I_T$,

$$\begin{aligned} &\text{tr}(S^{-1}X'\Upsilon\zeta\zeta'\Upsilon'XS^{-1}) \\ &= \sigma^{-2} \text{tr}(S^{-1}X'\Upsilon\Upsilon'X\eta\delta\delta'\eta'X'\Upsilon'XS^{-1}) \\ &= \sigma^{-2} \delta'\eta'\eta\delta. \end{aligned}$$

Similarly, when $\Omega = I_T$, we have

$$\begin{aligned} &\text{tr}(S^{-1}X'\Upsilon\bar{I}_J B \bar{I}_J \Upsilon'XS^{-1}) \\ &= \text{tr}(S^{-1}X'\Upsilon\Upsilon'XCX'\Upsilon'XS^{-1}) \\ &= \text{tr}(C) \end{aligned}$$

and

$$\begin{aligned} \delta'\eta'CX'\Omega^{1/2}\Upsilon P_3\zeta &= \delta'\eta'CX'\Upsilon\Upsilon'X\eta\delta h(2,0) \\ &= \delta'\eta'\eta\delta h(2,0). \end{aligned}$$

Therefore

$$\begin{aligned} \rho(\hat{\beta},\beta) &= \sigma^2\text{tr}(S^{-1}) - \sigma^2\text{tr}(C)h(2,0) \\ &\quad - \delta'\eta'\eta\delta\big(h(4,0)-h(2,0)\big) \end{aligned}$$

when $\Omega = I_T$. □

Appendix 4B

The Regressor Data

As the bias, risk and power functions are data dependent, a number of real and artificial regressor series have been chosen to evaluate the test and estimator properties. The artificial regressor series are,

- a) Random variables formed through the application of an AR(1) process (autocorrelation coefficient = 0.5) to standard normal data
- b) Log normal data, based on the standard normal distribution
- c) Exponentially trended data (16.182% increase per period)
- d) Standard normal data
- e) Linearly trended data
- f) Uniformly [0,1] distributed data.

For the heteroscedastic models the real regressor series chosen were,

- g) Australian real GDP, (annual, 1960-89)
- h) Australian money supply, (quarterly, 1960q1-89q4)
- i) New Zealand real GDP index, (quarterly 1972q2-90q2)*
- j) Australian\$ - U.S.\$ spot rate, (quarterly, 1960q1-90q1).

For the autoregressive and moving average models the real regressor series chosen were,

- k) Australian CPI, (quarterly 1960q1-90q2)
- l) Australian real retail trade, (quarterly 1960q1-90q1)
- m) Australian trade balance, (quarterly 1960q1-90q2).

All of the real regressor series are taken from the N.Z. Statistics Dept, I.N.F.O.S. database (source: O.E.C.D.), except *, whose source is the N.Z. Statistics Dept.

CHAPTER FIVE

PRE-TEST ESTIMATION OF THE COEFFICIENT AND PREDICTOR VECTORS IN A MODEL WITH A MIS-SPECIFIED ERROR COVARIANCE MATRIX AND EXCLUDED RELEVANT REGRESSORS

5.1 Introduction

In the previous chapter we considered the effects of an uncorrected autoregressive or heteroscedastic process in the error term on the risks of the OLS, RLS and PT estimators in the context of a model that is correctly specified in other respects. However, given that it is argued that such processes may arise partly through the exclusion of relevant regressors, it is, perhaps, unrealistic to consider such a process in isolation. In applied situations, regression models may frequently be mis-specified as a result of an unsound theoretical basis for the model, unobservable or incorrectly measured data, or over-simplification (among other reasons).

It is this problem that we address in this chapter, deriving the properties of the OLS and RLS estimators and the Wald statistic (for exact linear restrictions) in the context of a mis-specification of the error covariance matrix in conjunction with a mis-specification of the design matrix caused by excluded relevant regressors in Section 5.2, and the PTE bias and risk functions in Section 5.3. Details of the specific models used to evaluate these formulae numerically, and a summary of the results of the evaluations, are given in Section 5.4. Section 5.5 concludes the chapter.

5.2 Properties of the Component Estimators and the Test Statistic

Consider the linear regression model described by

$$y = X\beta + Z\beta_z + \varepsilon \quad ; \quad \varepsilon \sim N(0, \sigma^2 \Omega)$$

where, as usual, y is a $T \times 1$ vector of observations on the dependent variable, X and Z are non-stochastic matrices of size $T \times K$ and $T \times K_z$ respectively with $\text{Rank}(X) = K$ and $\text{Rank}(Z) = K_z$, β and β_z are unknown coefficient vectors of size $K \times 1$ and $K_z \times 1$ respectively and ε is a $T \times 1$ error term.

Assume that the fitted model is

$$y = X\beta + \mu$$

where μ is assumed to be $N(0, \sigma^2 I_T)$ when in fact it is $N(Z\beta_z, \sigma^2 \Omega)$ distributed. There are J non-stochastic restrictions to be tested and these are described by $H_0: R\beta = r$ vs $H_A: R\beta \neq r$. The validity of these restrictions is tested using the usual Wald statistic

$$u = \frac{(Rb-r)'(RS^{-1}R')^{-1}(Rb-r)(T-K)}{(y-Xb)'(y-Xb)J}$$

where $S = X'X$ and b is the OLS estimator of β , $b = S^{-1}X'y$. If the restrictions cannot be rejected, the model is estimated using the RLS estimator, $b^* = b + S^{-1}R'(RS^{-1}R')^{-1}(r-Rb)$. Otherwise OLS is used. Therefore the PTE can be written

$$\hat{\beta} = \begin{cases} b^* & \text{if } u < c \\ b & \text{if } u \geq c \end{cases}$$

where the critical value, c , is such that $\Pr.(F_{(J, \nu)} > c) = \alpha\%$, the nominal test size. Thus, the estimators are the classical restricted, unrestricted and pre-test estimators whose properties when applied in a well-specified model, are discussed in Chapter Two, and the choice of test statistic also reflects the fact the researcher is unaware that the model is mis-specified.

In this case it is straightforward to show that the bias and risk, under quadratic loss, of the unrestricted estimator are

$$B(b, \beta) = S^{-1}X'\xi$$

and

$$\rho(b, \beta) = \sigma^2 \text{tr}(S^{-1}X'\Omega XS^{-1}) + \xi'XS^{-2}X'\xi$$

where $\xi = Z\beta_z$. The bias and risk of the unrestricted predictor, Xb , are

$$B(Xb, E(y)) = E(XS^{-1}X' \varepsilon - M\xi) = -M\xi$$

and

$$\rho(Xb, \xi(y)) = \sigma^2 \text{tr}(X' \Omega X S^{-1}) + \xi' M \xi$$

respectively, where¹ $M = (I_T - X S^{-1} X')$.

Similarly, the bias and risk of the restricted estimator are

$$B(b^*, \beta) = -\eta \delta - (C - S^{-1})X' \xi$$

and

$$\begin{aligned} \rho(b^*, \beta) = & \delta' \eta' \eta \delta + 2\delta' \eta' (C - S^{-1})X' \xi + \xi' X(C - S^{-1})^2 X' \xi \\ & + \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1} - 2CX' \Omega X S^{-1} + CX' \Omega X C) \end{aligned}$$

respectively, where $\delta = R\beta - r$, $C = S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$ and $\eta = S^{-1}R'(RS^{-1}R')^{-1}$. The restricted predictor bias and risk are

$$B(Xb^*, E(y)) = -X\eta\delta - M\xi - X C X' \xi$$

and

$$\begin{aligned} \rho(Xb^*, E(y)) = & \sigma^2 \text{tr}(X' \Omega X (S^{-1} - C)) + \xi' M \xi \\ & + (\xi' X S^{-1} R' + \delta')(R S^{-1} R')^{-1}(\delta + R S^{-1} X' \xi) \end{aligned}$$

respectively. The unrestricted and restricted predictor risk functions collapse down to those given in Mittelhammer (1984) for the case where $\Omega = I_T$.

Now consider the test statistic, u . This statistic can be written as a ratio of quadratic forms in a Normal random vector:

¹

Note that the predictive risk of the unrestricted estimator differs from the estimator risk in the case of orthonormal regressors. This is in contrast to the case of a well-specified regression model in which they are the same.

$$u = \frac{\bar{\epsilon}' X C X' \bar{\epsilon} (T-K)}{\bar{\epsilon}' M \bar{\epsilon} (J)}$$

where $\bar{\epsilon} = (\epsilon + \xi + X\eta\delta)$ and, therefore, applying the method of Koerts and Abrahamse (1969), as outlined in Chapter Four, Section 4.2, we have

Theorem 5.1.

Given the assumptions of the model

$$\text{Pr.}(u < c) = \text{Pr.}(z' \Lambda z < 0)$$

$$= \text{Pr.} \left(\sum_{i=1}^T \lambda_i \chi_{(1; \theta_i)}^2 < 0 \right)$$

where $z = \sigma^{-1} \Gamma' \Omega^{-1/2} (X\eta\delta + \xi + \epsilon)$, $\Lambda = \text{diag}(\lambda_i)$ with Γ the eigenvector matrix, v_i the i 'th eigenvector and λ_i the i 'th eigenvalue of the matrix $\Phi \equiv \sigma^2 \Omega^{1/2} \left(X C X' - \frac{c J}{(T-K)} M \right) \Omega^{1/2}$. The non-centrality parameters of the $\chi_{(1; \theta_i)}^2$ statistics are $\theta_i = \frac{1}{2\sigma^2} \left(v_i' \Omega^{-1/2} (X\eta\delta + \xi) \right)^2$. Note that z is $N(\zeta, I_T)$ distributed with $\zeta \equiv \sigma^{-1} \left(\Gamma \Omega^{-1/2} (X\eta\delta + \xi) \right)$.

Proof.

The proof of this theorem is substantially the same as the proof of Theorem 4.1.

Corollary 5.1

When $\Omega = I_T$ the cumulative distribution function of the test statistic collapses down to the c.d.f. of a $F''_{(J, \nu; \lambda_1, \lambda_2)}$ random variable, with $\lambda_1 = \frac{1}{2\sigma^2} (RS^{-1} X' \xi + \delta)' (RS^{-1} R')^{-1} (RS^{-1} X' \xi + \delta)$ and $\lambda_2 = \frac{1}{2\sigma^2} \xi' M \xi$, as in Mittelhammer (1984).

Proof.

See Appendix 5A.

In general, however, the test statistic will not be distributed as a doubly non-central F random variable.

As in Chapter Four, we define a scalar level of hypothesis error in the regression model as

$$\theta \equiv \sum_{(1:\lambda_1 > 0)} \theta_1 = \sum_{(1:\lambda_1 > 0)} \frac{1}{2\sigma^2} \left(v_1' \Omega^{-1/2} (X\eta\delta + \xi) \right)^2$$

in this case.

As the level of hypothesis error, δ , increases the value of θ increases without bound for a given level of regressor mis-specification error, ξ . Note, however, that even if the null hypothesis is correct, θ will be non-zero if the model is mis-specified with respect to the regressors.

Following Mittelhammer (1984) we define a scalar measure of the degree of mis-specification in the regressor matrix as² $\lambda_2 = \frac{1}{2\sigma^2} \xi' M \xi$.

5.3 The Bias and Risk of the Pre-Test Predictor and the Pre-Test Estimator

The pre-test predictor and estimator are defined as

$$X\hat{\beta} = \begin{cases} Xb & \text{if } u \geq c \\ Xb^* & \text{if } u < c \end{cases} \quad \text{and} \quad \hat{\beta} = \begin{cases} b & \text{if } u \geq c \\ b^* & \text{if } u < c \end{cases}$$

respectively.

Theorem 5.2

a) The bias and risk of the pre-test predictor are given by

² Although it would be possible to define a scalar measure of regressor mis-specification as $\hat{\theta} = \sum_{(1:\lambda_1 < 0)} \theta_1$, where $\hat{\theta} = \lambda_2$ when $\Omega = I_T$, in practice this is less than ideal because varying the level of hypothesis error, δ , with ξ fixed will cause the value of $\hat{\theta}$ to vary when $\Omega \neq I_T$. Although it is possible to fix $\hat{\theta}$ this would entail varying ξ with δ and hence $\hat{\theta}$ would not, in fact, be measuring only the effect of regressor mis-specification.

$$B(\hat{X}\beta, E(y)) = -M\xi - XCX'\Omega^{1/2}\Gamma P_3 Y'\Omega^{-1/2}(X\eta\delta + \xi)$$

and

$$\begin{aligned} \rho(\hat{X}\beta, E(y)) &= \sigma^2 \text{tr}(X'\Omega XS^{-1}) + \xi' M \xi - \sigma^2 \text{tr}(XCX'\Omega^{1/2}\Gamma BY'\Omega^{1/2}) \\ &\quad + 2(X\eta\delta + \xi)' XCX'\Omega^{1/2}\Gamma P_3 Y'\Omega^{-1/2}(X\eta\delta + \xi) \end{aligned}$$

respectively, where P_3 and B are as defined above in Theorem 4.2, given the appropriate definition of $\zeta = E(z) = \sigma^{-1}Y'\Omega^{-1/2}(X\eta\delta + \xi)$ in this case.

b) The bias and risk of the pre-test estimator are given by

$$B(\hat{\beta}, \beta) = S^{-1}X'\xi - CX'\Omega^{1/2}\Gamma P_3 Y'\Omega^{-1/2}(X\eta\delta + \xi)$$

and

$$\begin{aligned} \rho(\hat{\beta}, \beta) &= \xi' XS^{-2}X'\xi + \sigma^2 \text{tr}(S^{-1}X'\Omega XS^{-1}) \\ &\quad - 2\sigma^2 \text{tr}(CX'\Omega^{1/2}\Gamma BY'\Omega^{1/2}XS^{-1}) + 2\delta'\eta' CX'\Omega^{1/2}\Gamma P_3 Y'\Omega^{-1/2}(X\eta\delta + \xi) \\ &\quad + \sigma^2 \text{tr}(CX'\Omega^{1/2}\Gamma BY'\Omega^{1/2}XC) \end{aligned}$$

respectively.

Proof.

See Appendix 5A.

As in the case where the regressor matrix is well-specified, the bias and risk of the pre-test predictor and estimator depend on the form of the data generating process, the form of the restrictions and the critical value chosen for the pre-test. Similarly we have:

Corollary 5.2.

As the pre-test critical value, c , tends towards zero (∞), the bias and risk of the PT predictor and estimator tend towards the bias and risk of the unrestricted (restricted) estimator.

Proof.

See Appendix 5A.

Using Lemma 4.2 it is straightforward to show that the PT predictor and the estimator bias and risk functions collapse down to the corresponding functions for a linear model with a scalar covariance matrix when $\Omega = I_T$.

Corollary 5.3.

When $\Omega = I_T$

a) the PT predictor bias and risk functions are given by

$$B(\hat{X}\hat{\beta}, E(y)) = -(M + XCX'h'(2,0))\xi - X\eta\delta h'(2,0)$$

and³

$$\rho(\hat{X}\hat{\beta}, E(y)) = (K + 2\lambda_2 + (4\lambda_1 - J)h'(2,0) - 2\lambda_1 h'(4,0))\sigma^2.$$

b) the PT estimator bias and risk functions are given by

$$B(\hat{\beta}, \beta) = (S'X' - CX'h'(2,0))\xi - \eta\delta h'(2,0)$$

and

$$\begin{aligned} \rho(\hat{\beta}, \beta) = & \sigma^2 \text{tr}(S^{-1}) + \xi XS^{-2}X'\xi + 2\sigma^2(\delta'\eta'\eta\delta + \delta'\eta'CX'\xi)h'(2,0) \\ & - \sigma^2 \text{tr}(C)h'(2,0) \\ & - (\delta' + \xi XS^{-1}R')\eta'\eta(RS^{-1}X'\xi + \delta)h'(4,0) \\ & - 2\xi'\bar{I}_K X' S^{-2} X Y' \bar{I}_J \xi h'(0,0) \\ & - 2\xi'\bar{I}_T X' S^{-2} X Y' \bar{I}_J \xi h'(0,2), \end{aligned}$$

where

$$h'(i,j) = \Pr. \left(\frac{\chi_{(J;\lambda_1)}^2}{\chi_{(T-K;\lambda_2)}^2} < \frac{cJ}{T-K} \right),$$

$$\bar{I}_J = \begin{bmatrix} I_J & 0 \\ 0 & 0 \end{bmatrix}, \bar{I}_K = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_{K-J} & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \bar{I}_T = \begin{bmatrix} 0 & 0 \\ 0 & I_{T-K} \end{bmatrix}.$$

Note that in a model with a well-specified error covariance matrix, the unrestricted, restricted and pre-test predictor risk functions are

³ This expression is equivalent to the one derived by Mittelhammer (1984) for the PT predictor in a model with a mis-specified regressor matrix and a well specified error covariance matrix.

independent of the data, other than through T , K , J and the non-centrality parameters λ_1 and λ_2 . This is not the case in a model in which the error covariance matrix is mis-specified. It is apparent that, as for the estimators for the coefficient vector, the predictor risks depend on the particular form of the regressor matrix and the restrictions in this situation. It is difficult, therefore, to make further observations regarding them without numerical evaluation. This has been done for both the coefficient and conditional expectation estimators using a FORTRAN program written by the author and the Davies' (1980) algorithm on a Vax 6340 computer. The particular models used and the results obtained are discussed in Section 5.4.

5.4 Numerical Evaluation

5.4.1 Numerical Considerations

The models used in evaluating the risk functions derived above can be written in the general form

$$y_t = x_t\beta + z_t\beta_z + u_t \quad ; \quad t = 1, \dots, T$$

where the errors follow an autoregressive AR(1), MA(1), or AR(4) process or one of a number of heteroscedastic processes as set out in Chapter Four, Section 4.3. The restrictions are written in the usual way as $H_0: R\beta = r$ vs $H_A: R\beta \neq r$, and the particular data series used are described in Appendix 4B.

Note, however, that in contrast to the situation in which the regressor matrix is well-specified, the restricted and pre-test estimators of the coefficient vector are not directly related to the non-centrality parameter, θ , even if there is, in fact, no mis-specification of the error covariance matrix (in which case we may write $\theta = \lambda_1 = \frac{1}{2\sigma^2}(\delta' + \xi'XS^{-1}R)(RS^{-1}R')^{-1}(RS^{-1}X'\xi + \delta)$). Let us consider the risk of the restricted estimator of the

coefficient vector in a model in which $J = 1$ (i.e. δ is a scalar and $\Omega = I_T$). In this case $\rho(b^*, \beta) = \delta' \eta' \eta \delta + 2\delta' \eta' (C-S^{-1})X' \xi + \xi' (C-S^{-1})^2 X' \xi + \sigma^2 \text{tr}(S^{-1}-C)$.

It is clear that the restricted estimator's risk is a quadratic in δ , as is the non-centrality parameter λ_1 . If $\xi = 0$, then $\rho(b^*, \beta)$ and λ_1 are linearly related, but they are not if $\xi \neq 0$. This arises because there are two values of δ which will give rise to each particular value of λ_1 , which will not, in general, give rise to the same value of $\rho(b^*, \beta)$. As we would expect, the pre-test estimator's risk is not related on a one-to-one basis with the non-centrality parameter either, with two possible values of $\rho(\hat{b}, \beta)$ being associated with each value of λ_1 . This problem will also exist in the case where $\Omega \neq I_T$ and, as a consequence of this, we will consider the risks of the coefficient vector estimators to be functions of δ rather than⁴ of θ .

As noted by Mittelhammer (1984) and others, our definition of the non-centrality parameter, in the case of $\Omega = I_T$ implies that, in general, λ_1 will not equal 0 even if the prior information is valid ($\delta = 0$). It will also be true that when $\Omega \neq I_T$, θ will not, in general, be equal to 0 if the prior information is valid. In fact, as θ is a quadratic in δ , it may be the case that θ will not be equal to zero anywhere in the parameter space. It is clear that, when $\xi = 0$, the minimum value of θ , denoted $\bar{\theta}$, equals 0 regardless of the data set under consideration, however when $\xi \neq 0$, the value of $\bar{\theta}$ will vary, for a given value of λ_2 , with the particular form of

⁴ Note that the restricted predictor is directly related to λ_1 when $\Omega = I_T$ and hence, when the error covariance matrix is well specified, $\rho(Xb^*, E(y))$ is linear in λ_1 . In fact, if $\Omega = I_T$ the value of $\rho(Xb^*, E(y))$ may differ slightly depending on which root of θ we consider, qualitatively and quantitatively, however the difference is minor.

regressor matrix and restrictions considered. Therefore, although it is true that, when $\Omega = I_T$, the restricted predictor will have a lower risk than the unrestricted predictor when $\lambda_1 \leq \frac{J}{2}$ (see Mittelhammer (1984)) it may be the case that, for a particular data set, the minimum value that λ_1 can take is greater than $\frac{J}{2}$ and hence the restricted, and the pre-test predictors, are strictly dominated by the unrestricted predictor. It is straightforward to show that, in general, $\bar{\theta}$ is increasing in λ_2 .

Recall that $\Pr.(u < c) = \Pr.\left(\sum_{i=1}^T \lambda_i \chi_{(1;\theta_i)}^2 < 0\right)$. From the definition of θ it is apparent that this probability is increasing in θ , hence the test will have minimum power at the point $\theta = \bar{\theta}$. Let the value of θ under the null hypothesis be denoted θ^0 . As $\theta^0 > \bar{\theta}$, in general, when $\lambda_2 \neq 0$, we see that the pre-test may be biased, that is, the power of the test may not be greater than the true test size. Having made these general points we now turn to some numerical evaluations.

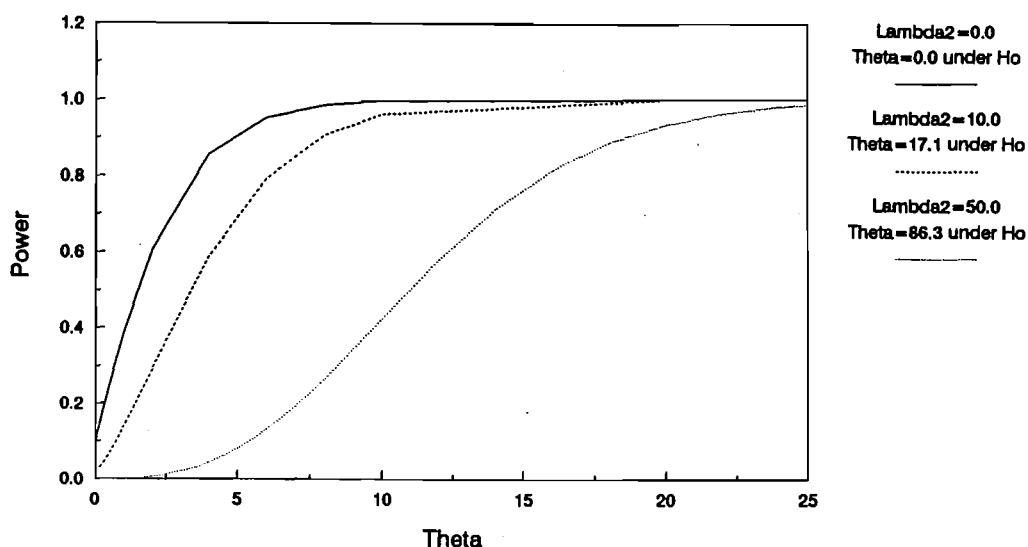
5.4.2 Numerical Results

Quantitatively the effect of excluded regressors on the estimators under consideration is to increase the risk of the estimators for each value of λ_2 (and δ). This is apparent from the bias and risk formulae derived above. The effect of the mis-specification of the error covariance matrix is also likely to increase the risks of each of the estimators. As in Chapter Four, however, there are instances in which the estimators' risks are slightly reduced by an autoregressive or heteroscedastic error process. Qualitatively the effects of the double mis-specification vary according to the estimators we apply and to the specific data we consider, although, as the level of regressor mis-specification increases, the additional effects of a mis-specified error covariance matrix may be qualitatively slight.

We will first consider the effect of the double mis-specification on the test power function, as this relates to both coefficient vector estimators and predictors. As we would expect, the power of the test falls, as λ_2 increases for a given value of θ . This occurs because as λ_2 increases the $\chi^2_{(1;\theta_1)}$ variables associated with negative eigenvalues in the c.d.f. of the test statistic, $\Pr\left(\sum_{i=1}^T \lambda_i \chi^2_{(1;\theta_1)} < 0\right)$ become increasingly non-central, whereas the $\chi^2_{(1;\theta_1)}$ r.v.'s associated with positive eigenvalues do not, as θ is fixed. Hence, the probability of the weighted sum of $\chi^2_{(1;\theta_1)}$ variables taking a value of less than zero (or, equivalently, the probability that the test statistic will lie below the chosen critical value), increases and the power of the test is reduced for fixed θ . However, the true size of the test will generally increase as the value of θ^0 increases as λ_2 increases. The test thus becomes increasingly biased as the regressor mis-specification increases, as is shown in Figure 5.1.

FIGURE 5.1: Pre-test Power Functions

Excluded Regressor, Scalar Error Covar. Matrix
Three regressors



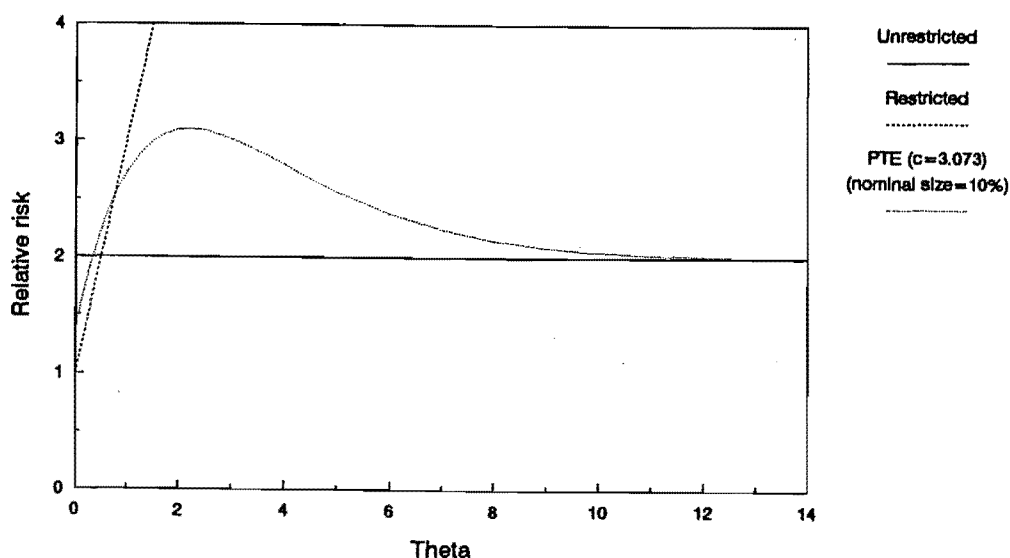
One restriction
Excluded regressor is exponential trend

Sample size = 18
Nominal size = 10%

The effect of this increase in θ^0 on the pre-test power function is qualitatively minor in terms of estimating the prediction vector, as Figures 5.2 and 5.3 show. This is because the restricted and pre-test estimators are related, not to δ the constraint specification error, but to θ and therefore the changing value of θ^0 will have no qualitative effect on these functions. This is not the case when we consider the coefficient vector estimators, as in Figures 5.4 and 5.5. Even for a relatively modest value of $\lambda_2 = 10$ the true size of the test is close to unity in this case, while the power of the test initially falls as δ decreases below zero. If the regressor mis-specification is severe enough it may well be the case that the only region in which the pre-test is likely to accept the restrictions is one in which the restricted estimator is dominated by the unrestricted estimator. An example of this is shown in Figure 5.6.

FIGURE 5.2: Predictor Risk Functions

Well specified model
Three regressors

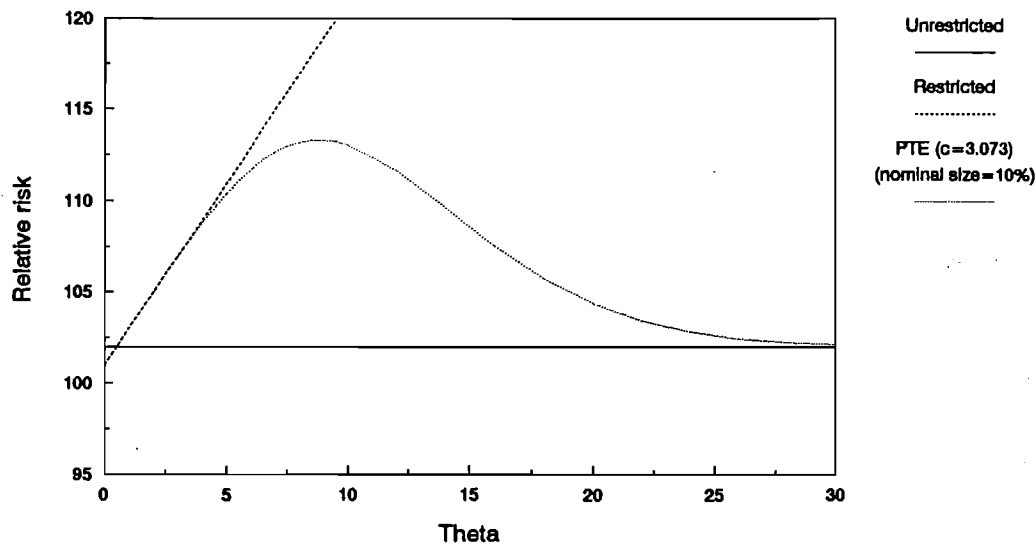


One restriction

Sample size = 18

FIGURE 5.3: Predictor Risk Functions

Excluded Regressor, Scalar Error Covar. Matrix
Regressors; constant, Australian CPI and trend

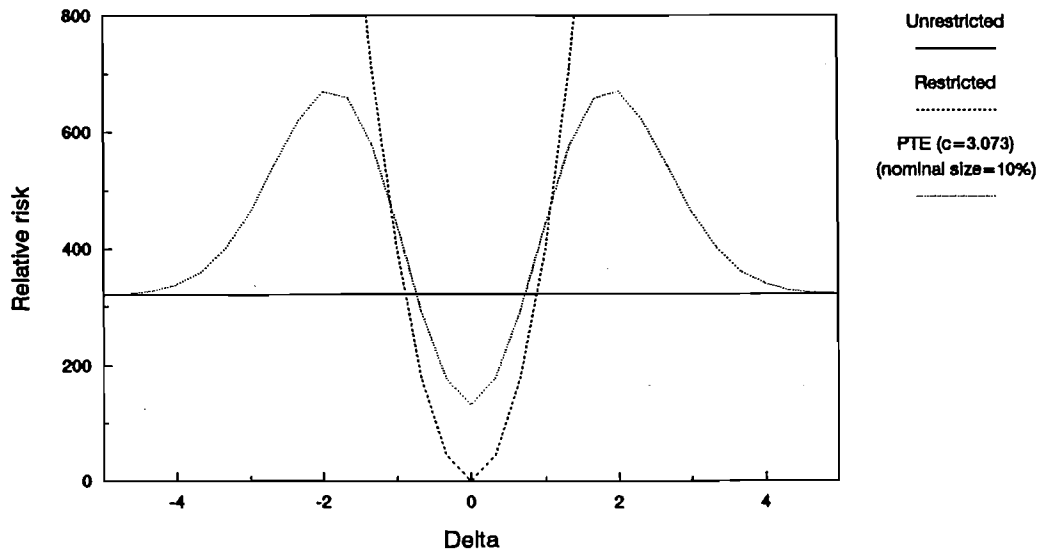


One restriction on Australian CPI
Excluded regressor is exponential trend

Sample size = 18
 $\lambda_2 = 10.0$

FIGURE 5.4: Coefficient Estimator Risk Functions

Well specified model
Regressors; constant, Australian CPI and trend

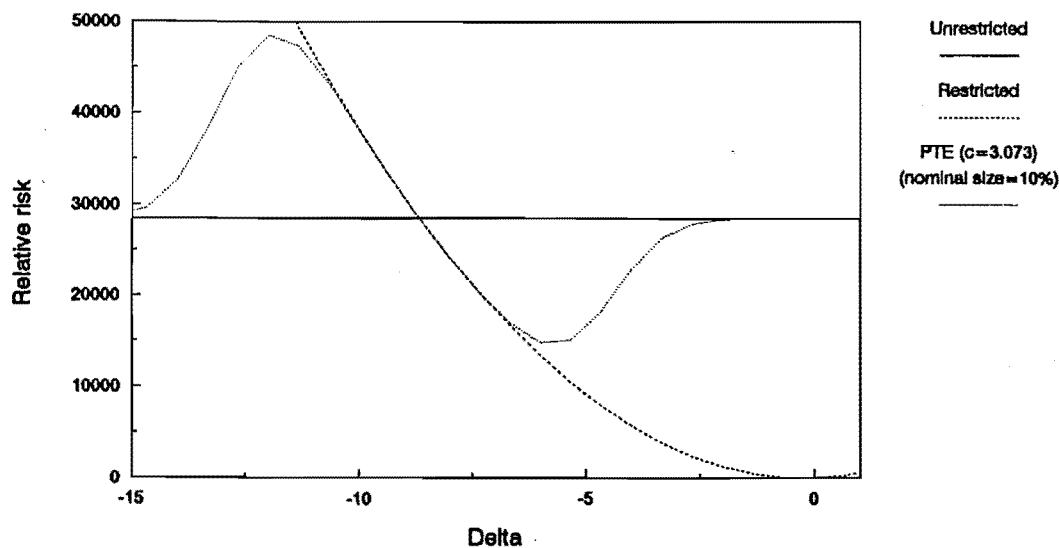


One restriction on Australian CPI

Sample size = 18

FIGURE 5.5: Coefficient Estimator Risk Functions

Excluded Regressor, Scalar Error Covar. Matrix
Regressors; constant, Australian CPI and trend

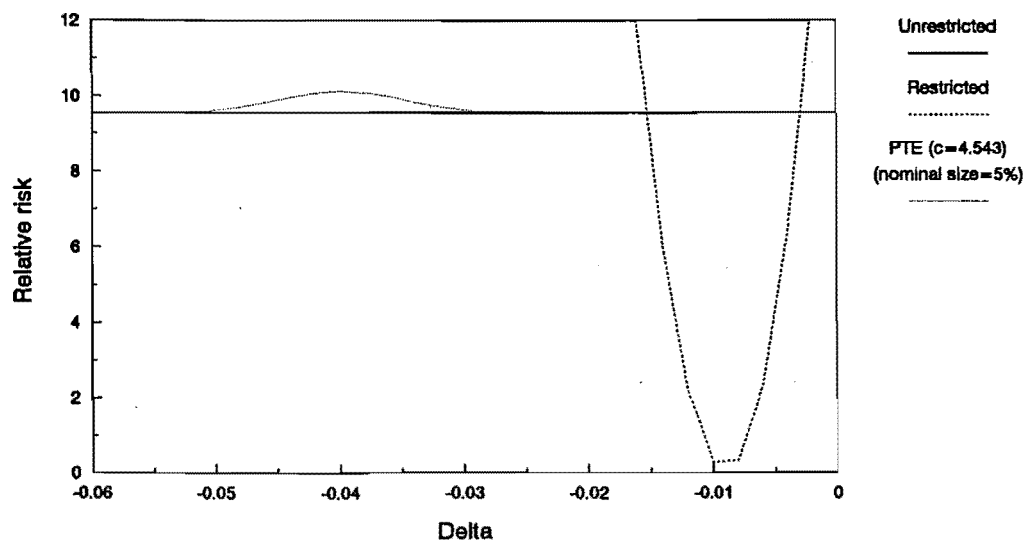


One restriction on Australian CPI
Excluded regressor is exponential trend

Sample size = 18
 $\text{Lambda2} = 10.0$

FIGURE 5.6: Coefficient Estimator Risk Functions

Excluded Regressor, Scalar Error Covar. Matrix
Regressors; constant, Aust. Retail Trade & trend



Two restrictions on Aust. Retail Trade and trend
Excluded regressor is exponential trend

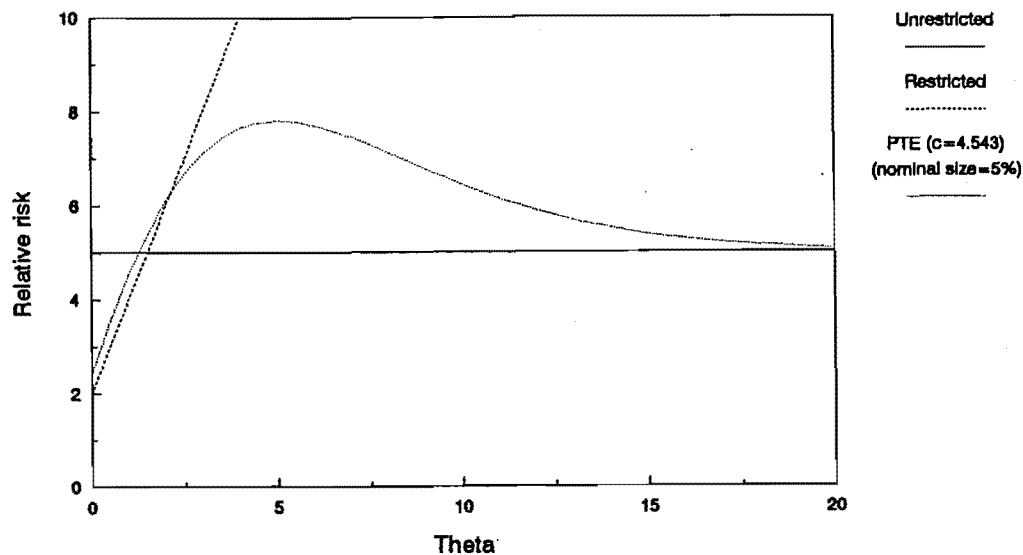
Sample size = 18
 $\text{Lambda2} = 10.0$

The effect of a mis-specified error covariance matrix on the pre-test power function is qualitatively the same for a given degree of regressor mis-specification regardless of what that degree is. As we saw in Chapter Four, the power of the test is generally distorted upwards in the case of positive MA(1) or AR(1) errors, particularly if the regressors whose coefficients are included in the restrictions are trended, and conversely for a negative MA(1) or AR(1) process. Similarly, if the errors are generated by an AR(4) process, the power of the test is reduced if the absolute value of ρ_4 increases. The effects of uncorrected heteroscedasticity are also qualitatively the same as those discussed in Chapter Four.

We now consider the direct effects of the mis-specification on the estimators, rather than the indirect effect through the pre-test power function. In the case of the prediction vector estimators the risk functions are independent of the data, other than through T , K , J , θ , λ_2 and $\bar{\theta}$ the minimum value of $\bar{\theta}$, when the error covariance matrix is scalar. However, the value of $\bar{\theta}$ varies with the particular data set under consideration. Recall that, if $\bar{\theta} > \frac{J}{2}$ the restricted and pre-test estimators will be strictly dominated by the unrestricted estimator, as shown in Figures 5.7 and 5.8. From our numerical results, it appears that $\bar{\theta}$ increases with the correlation between the regressors whose coefficients are included in the restrictions being tested, and decreases with the correlation between the included regressors whose coefficients are not being tested, and the excluded regressors. This makes sense intuitively as, if the excluded regressors were not correlated with the included regressors, we would expect the exclusions

FIGURE 5.7: Predictor Risk Functions

Well specified model
Five regressors

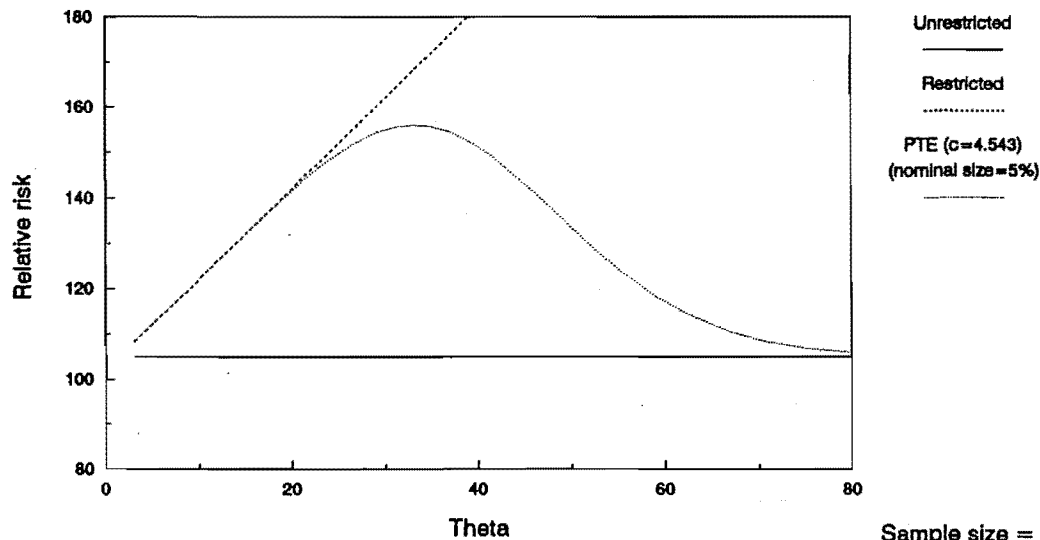


Three restrictions

Sample size = 18

FIGURE 5.8: Predictor Risk Functions

Excluded Regressor, Scalar Error Covar. Matrix
Regressors, const. uniform[0,1] r.v. & dummies



Three restrictions on seasonal dummies
Excluded regressor is $N(0,1)$ r.v.

Sample size = 18
 $\Lambda_2 = 10.0$
 $\Theta_{\text{bar}} = 3.069$

to have no effect.⁵ On the other hand, if the regressors included in the model are highly correlated with those that are excluded they may be acting as "proxy" variables thereby reducing the effects of the regressor mis-specification. $\bar{\theta}$ also increases with the degree of regressor mis-specification, λ_2 , and the number of restrictions being tested.

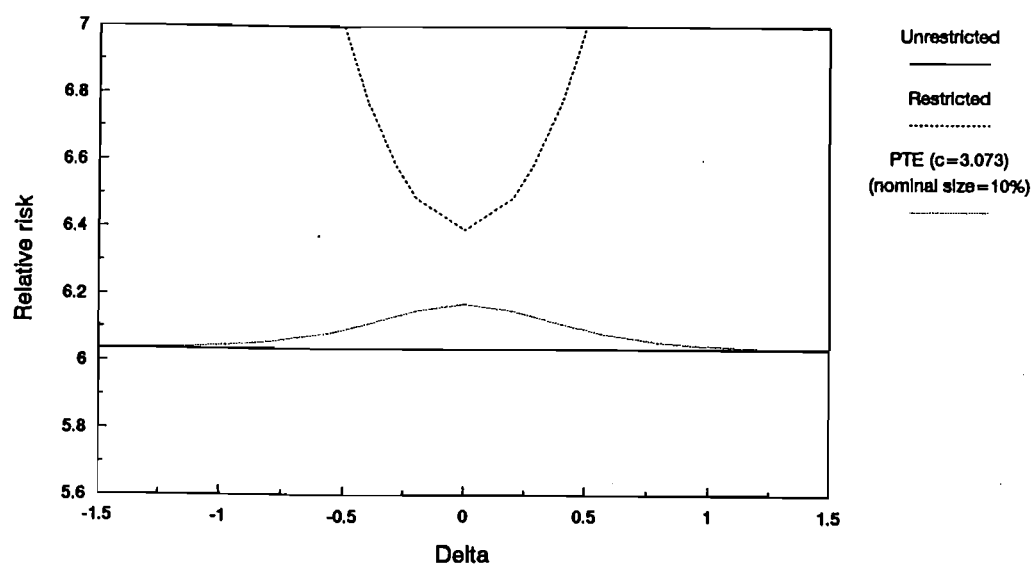
If, in addition to having a mis-specified regressor matrix, the model suffers from a mis-specified error covariance matrix the same general results appear to hold. If λ_2 is small in magnitude, ($\lambda_2 < 5-10$), it may be the case that an autoregressive or heteroscedastic process in the error term may offset, or accentuate, the effect of the excluded regressors on $\bar{\theta}$. From our numerical results it appears that the value of $\bar{\theta}$ may decrease as ρ_1 , ρ_4 or ρ increase. The effect of a heteroscedastic process depends on the particular process and the data being considered. However, if λ_2 is large, the presence of the non-scalar error covariance matrix appears to make little difference to the estimators' risks, qualitatively, other than the effect on the pre-test power function discussed above.

In the case of the coefficient vector estimators, an increase in the value of λ_2 generally has the effect of increasing the range, of δ , over which the restricted estimator has lower risk than the unrestricted estimator, as is shown in Figures 5.4 and 5.5. As we have seen, however, this may not mean that pre-testing becomes more attractive relative to the OLS estimator as, if λ_2 is large, the pre-test may reject the restrictions over the range of θ in which the RLS estimator dominates the OLS estimator.

⁵ Alternatively consider the case of testing for exclusion restrictions. If the regressors whose significance is being tested are highly correlated with the excluded regressors they may be serving as "proxy" variables (see McCallum (1972), Wickens (1972) or Ohtani (1983), among others) for the excluded variables and, therefore the restricted model, is more likely to be dominated by the unrestricted model the greater is the correlation.

FIGURE 5.9: Coefficient Estimator Risk Functions

No Excluded Regressors, AR(1) Errors, $\rho = 0.9$
Regressors; const., trend and r.v. with AR(1)

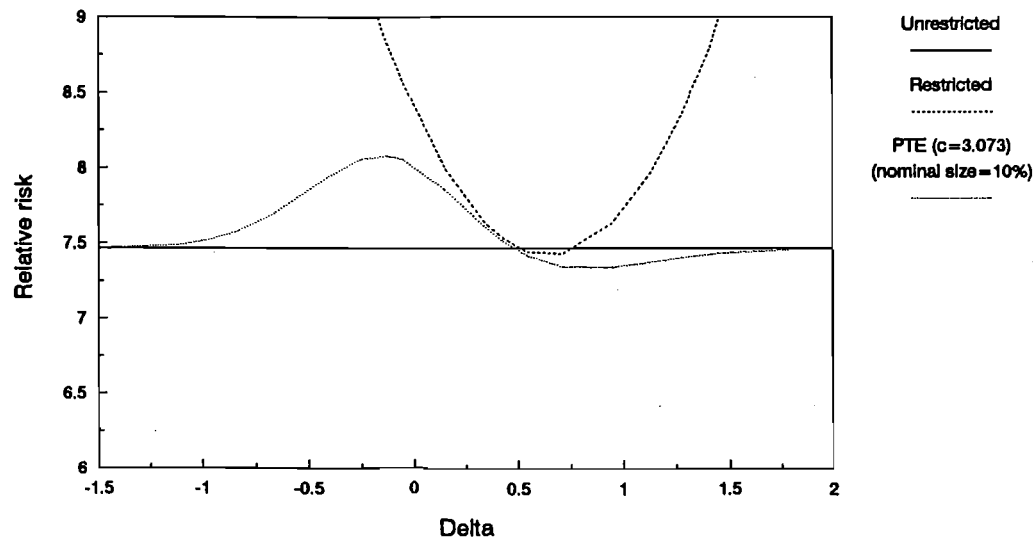


One restriction on $N(0,1)$ r.v. with AR(1)

Sample size = 18

FIGURE 5.10: Coefficient Estimator Risk Fns.

Excluded Regressor, AR(1) Errors, $\rho = 0.9$
Regressors; const., trend and r.v. with AR(1)



One restriction on $N(0,1)$ r.v. with AR(1)
Excluded regressor is exponential trend

Sample size = 18
 $\text{Lambda2} = 10.0$

When estimating either the coefficient or prediction vector, if λ_2 is sufficiently large the effect of the mis-specified error covariance matrix appears to be qualitatively minor, other than the effect on the pre-test's power function. Consider, for example, Figures 5.9 and 5.10. Even though the RLS and PT estimators are strictly dominated when $\lambda_2 = 0$ and $\rho_4 = 0.6$., as λ_2 increases the risk functions take on the shape typical of these estimators in a model that is only mis-specified through the exclusion of relevant regressors.

Other things being equal, an increase in the sample size leads to a reduction in estimator risk for each value of θ . In addition it leads to a reduction in the value of $\bar{\theta}$, thus increasing the range over which the restricted and pre-test predictors have lower risk than the unrestricted predictor, if such a range exists. An increase in sample size also reduces the distortion of the pre-test's power function caused by the regressor mis-specification, although it may not reduce the distortion caused by a mis-specified error covariance matrix, as we have seen in Chapter Four.

5.5 Conclusion

The practical implications of the double form of model mis-specification considered in this chapter vary depending on the particular estimators considered, the degree of regressor mis-specification and, to a lesser extent, the nature of the true error covariance matrix and the regressors. The effect of these mis-specifications on the pre-test's power function may be that the true size of the pre-test is close to unity, while the test may have quite a low power in some other parts of the parameter space. This may lead to a pre-test estimator of the coefficient vector that is strictly dominated by the corresponding unrestricted estimator.

Similarly, in the case of the unrestricted, restricted and pre-test predictors, the PTE and, in this case, the restricted estimator, may be strictly dominated by the unrestricted predictor, particularly if the sample size is small and a number of restrictions are being tested. One implication of this study is that it may be preferable, in terms of minimizing estimator risk, to retain a theoretically, irrelevant regressor in a regression model if it is thought that it may be correlated with a relevant but excluded regressor. That is to say that these results add further weight to the concept of using a "proxy" variable model if the data for one or more of the regressors in the theoretical model cannot be obtained. See McCallum (1972), Wickens (1972), Aigner (1974) and Ohtani (1983, 1985), among others, for further discussion.

From the particular case we consider in this chapter, it appears to be the case that a mis-specified error covariance matrix will make little difference, qualitatively, to the properties of the estimators unless the regressor matrix is, or is close to being, well specified. Other things being equal, the greater the value of λ_2 the less effect the mis-specification of the error covariance matrix has, other than on the pre-test's power function.

Appendix 5A

Proof of Corollary 5.1.

By Lemma 4.1, the matrices $\left(XCX' - \frac{cJ}{T-K}M\right)$, XCX' and M have a common eigenvector matrix \bar{I} . When $\Omega = I_T$, $\Phi = \sigma^2 \Omega^{1/2} \left(XCX' - \frac{cJ}{T-K}M\right) \Omega^{1/2}$ also shares this eigenvector matrix and has eigenvalues

$$\begin{aligned} \lambda_i &= \sigma^2 & ; & \quad i = 1, \dots, J \\ \lambda_i &= 0 & ; & \quad i = J+1, \dots, K \\ \lambda_i &= -\sigma^2 \frac{cJ}{T-K} & ; & \quad i = K+1, \dots, T. \end{aligned}$$

$$\begin{aligned} \text{Therefore, when } \Omega = I_T, \Pr(u < c) &= \Pr\left(\sum_{i=1}^T \lambda_i \chi_{(1;\theta_i)}^2 < 0\right) \\ &= \Pr\left(\sum_{i=1}^J \sigma^2 \chi_{(1;\theta_i)}^2 + \sum_{i=J+1}^K 0 \chi_{(1;\theta_i)}^2 + \sum_{i=K+1}^T -\frac{\sigma^2 cJ}{T-K} \chi_{(1;\theta_i)}^2 < 0\right). \end{aligned}$$

Now, $\theta_i = \frac{1}{2\sigma^2} (v_i' (\xi + X\eta\delta))^2$ and therefore, when $\Omega = I_T$

$$\sum_{i=1}^J \theta_i = \frac{1}{2\sigma^2} \sum_{i=1}^T ((\xi + X\eta\delta)' X C X' v_i)^2$$

as

$$X C X' v_i = \begin{cases} v_i & ; \quad i = 1, \dots, J \\ 0 & ; \quad i = J+1, \dots, T, \text{ when } \Omega = I_T \end{cases}$$

by Lemma 4.2.

Hence, when $\Omega = I_T$,

$$\begin{aligned} \sum_{i=1}^J \theta_i &= \frac{1}{2\sigma^2} (\xi' + \delta' \eta' X') X C X' Y Y' X C X' (\xi + X\eta\delta) \\ &= \frac{1}{2\sigma^2} (\xi + \delta' (R S^{-1} R')^{-1} R S^{-1} X') X S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X' \\ &\quad X S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X' (\xi + X S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X') \end{aligned}$$

$$= \frac{1}{2\sigma^2} (\xi X S^{-1} R' + \delta') (R S^{-1} R')^{-1} (R S^{-1} X' \xi + \delta) = \lambda_1.$$

Similarly, when $\Omega = I_T$,

$$Mv_i = \begin{cases} 0 & ; \quad i = 1, \dots, K \\ v_i & ; \quad i = K+1, \dots, T, \end{cases}$$

and

$$\begin{aligned} \sum_{i=K+1}^T \theta_i &= \frac{1}{2\sigma^2} (\xi' + \delta' \eta' X') M Y Y' (\xi + X \eta \delta) \\ &= \frac{1}{2\sigma^2} \xi' M \xi = \lambda_2. \end{aligned}$$

Therefore, when $\Omega = I_T$, $\Pr.(u < c) = \Pr. \left(\chi_{(J; \lambda_1)}^{2,} - \frac{cJ}{T-K} \chi_{(v; \lambda_2)}^{2,} < 0 \right)$. \square

Proof of Theorem 5.2.

First note that, given appropriate definitions of z and b , Lemma 4A.1 holds, i.e.

$$E[z \Psi(z' \Lambda z)] = P_3 \zeta$$

and

$$E[zz' \Psi(z' \Lambda z)] = B.$$

$$\begin{aligned} \text{a) } \hat{\beta} &= \beta + S^{-1} X' (\varepsilon + \xi) - S^{-1} R' (R S^{-1} R')^{-1} (r - R\beta - S^{-1} X' (\varepsilon + \xi)) \\ &= \beta + S^{-1} X' (\varepsilon + \xi) - \sigma C X' \Omega^{1/2} \Upsilon z \Psi(z' \Lambda z) \end{aligned}$$

where

$$\Psi(x) = \begin{cases} 0 & x \geq 0 \\ 1 & x < 0 \end{cases}.$$

Therefore,

$$\begin{aligned} B(X\hat{\beta}, E(y)) &= E[X\beta + XS^{-1}X'(\varepsilon + \xi) - \sigma CX' \Omega^{1/2} \Upsilon z \Psi(z' \Lambda z)] - X\beta - \xi \\ &= XS^{-1}X' \xi - \xi - CX' \Omega^{1/2} \Upsilon P_3 \Upsilon' \Omega^{-1/2} (X\eta \delta + \xi) \end{aligned}$$

$$= -M\xi - XCX'\Omega^{1/2}\Upsilon P_3\Upsilon'\Omega^{1/2}(X\eta\delta+\xi)$$

by Lemma 4A.1.

Similarly,

$$\begin{aligned}\rho(\hat{X}\beta, E(y)) &= E[(\hat{X}\beta - X\beta - \xi)'(\hat{X}\beta - X\beta - \xi)] \\ &= E[(XS^{-1}X'\varepsilon - M\xi - \sigma XCX'\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z))'(XS^{-1}X'\varepsilon - M\xi \\ &\quad - \sigma XCX'\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z))] \\ &= E[\varepsilon'XS^{-1}X'XS^{-1}X'\varepsilon - 2\varepsilon'XS^{-1}X'M\xi - 2\sigma\varepsilon'XS^{-1}X'XCX'\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z) \\ &\quad + \xi'M\xi' + 2\sigma\xi'MXCX'\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z) \\ &\quad + \sigma^2Z'\Upsilon'\Omega^{1/2}XCX'\Psi(z'\Lambda_Z)XCX'\Omega^{1/2}\Upsilon_Z] \\ &= \sigma^2\text{tr}(X'\Omega XS^{-1}) + \xi'M\xi - \sigma^2\text{tr}(XCX'\Omega^{1/2}\Upsilon_B\Upsilon'\Omega^{1/2}) \\ &\quad + 2(X\eta\delta+\xi)'XCX'\Omega^{1/2}\Upsilon P_3\Upsilon'\Omega^{1/2}(X\eta\delta+\xi)\end{aligned}$$

by Lemma 4A.1.

$$\begin{aligned}\text{b)} \quad B(\hat{\beta}, \beta) &= E(\beta + \Omega^{1/2}X'(\varepsilon + \xi) - \sigma CX'\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z) - \beta) \\ &= S^{-1}X'\xi - CX'\Omega^{1/2}\Upsilon P_3\Upsilon'\Omega^{1/2}(X\eta\delta+\xi)\end{aligned}$$

by Lemma 4A.1, and

$$\begin{aligned}\rho(\hat{\beta}, \beta) &= E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] \\ &= E\left[(\varepsilon'XS^{-1} + \xi'XS^{-1} - \sigma Z'\Upsilon'\Omega^{1/2}XC\Psi\Omega^{1/2}XC\Psi(z'\Lambda_Z))\right. \\ &\quad \left.(S^{-1}X'\varepsilon + S^{-1}X'\xi - \sigma CX'\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z))\right] \\ &= E[\varepsilon'XS^{-1}S^{-1}X'\varepsilon + 2\varepsilon'XS^{-1}S^{-1}X'\xi - 2\sigma\varepsilon'XS^{-1}CX'\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z) \\ &\quad + \xi'XS^{-1}S^{-1}X'\xi - 2\sigma\xi'XS^{-1}CX'\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z) \\ &\quad + \sigma^2Z'\Upsilon'\Omega^{1/2}XC\Psi\Omega^{1/2}\Upsilon_Z\Psi(z'\Lambda_Z)] \\ &= \sigma^2\text{tr}(S^{-1}X'\Omega XS^{-1}) - 2\sigma^2\text{tr}(CX'\Omega^{1/2}\Upsilon_B\Upsilon'\Omega^{1/2}XS^{-1}) \\ &\quad + 2(X\eta\delta+\xi)'XS^{-1}CX'\Omega^{1/2}\Upsilon P_3\Upsilon'\Omega^{1/2}(X\eta\delta+\xi)\end{aligned}$$

$$\begin{aligned}
& + \xi' X S^{-2} X' \xi - 2\xi' X S^{-1} C X' \Omega^{1/2} \Gamma P_3 \Gamma' \Omega^{-1/2} (X\eta\delta + \xi) \\
& + \sigma^2 \text{tr}(C X' \Omega^{1/2} \Gamma B \Gamma' \Omega^{1/2} C X)
\end{aligned}$$

by Lemma 4A.1. Hence,

$$\begin{aligned}
\rho(\hat{\beta}, \beta) &= \sigma^2 \text{tr}(S^{-1} X' \Omega X S^{-1}) + \xi' X S^{-2} X' \xi + 2\delta' \eta' C X' \Omega^{1/2} \Gamma P_3 \Gamma' \Omega^{-1/2} (X\eta\delta + \xi) \\
&- \sigma^2 \text{tr}(C X' \Omega^{1/2} \Gamma B \Gamma' \Omega^{-1/2} X S^{-1}) + \sigma^2 (C X' \Omega^{1/2} \Gamma B \Gamma' \Omega^{-1/2} X C). \quad \square
\end{aligned}$$

Proof of Corollary 5.2.

As $c \rightarrow 0(\omega)$, the negative eigenvalues of Φ tend toward $0(\omega)$ and hence the P_{31} , P_{51} and $P_{31,3j}$ values tend towards 0 (unity) as shown above in the proof of Corollary 4.2, Appendix 4. In the former case, the proof is trivial; in the latter case we have

$$\begin{aligned}
\lim_{c \rightarrow \infty} B(\hat{X}\hat{\beta}, E(y)) &= -M\xi - X C X' \Omega^{1/2} \Gamma I_T \Gamma' \Omega^{-1/2} (X\eta\delta + \xi) \\
&= -M\xi - X S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X' (X S^{-1} R' (R S^{-1} R')^{-1} \delta + \xi) \\
&= (-M - X C X') \xi - X \eta \sigma \\
&= B(Xb^*, E(y)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\lim_{c \rightarrow \infty} \rho(\hat{X}\hat{\beta}, E(y)) &= \sigma^2 \text{tr}(X' \Omega X S^{-1}) + \xi' M \xi - \sigma^2 \text{tr}(X C X' \Omega^{1/2} \Gamma (I_T + \xi \xi') \Gamma' \Omega^{1/2}) \\
&+ 2(X\eta\delta + \xi)' + X C X' \Omega^{1/2} \Gamma I_T \Gamma' \Omega^{-1/2} (X\eta\delta + \xi) \\
&= \sigma^2 \text{tr}(X' \Omega X S^{-1}) + \xi' M \xi - \sigma^2 \text{tr}(X C X' \Omega) \\
&- (\delta' \eta' X' + \xi') \Omega^{-1/2} \Gamma \Gamma' \Omega^{1/2} X C X' \Omega^{1/2} \Gamma \Gamma' \Omega^{-1/2} (X\eta\delta + \xi) \\
&+ 2(\delta' \eta' X' + \xi') X C X' (X\eta\delta + \xi) \\
&= \rho(Xb^*, E(y)).
\end{aligned}$$

Also,

$$\begin{aligned}
\lim_{c \rightarrow \infty} B(\hat{\beta}, \beta) &= S^{-1} X' \xi - C X' \Omega^{1/2} \Gamma I_T \Gamma' \Omega^{-1/2} (X\eta\delta + \xi) \\
&= S^{-1} X' \xi - C X' \xi - S^{-1} R' (R S^{-1} R')^{-1} R S^{-1} X' X S^{-1} R' (R S^{-1} R')^{-1} \delta
\end{aligned}$$

$$= B(b^*, \beta)$$

and

$$\begin{aligned}
\text{Limit}_{c \rightarrow \infty} \rho(\hat{\beta}, \beta) &= \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1}) + \xi' X S^{-2} X' \xi \\
&\quad + 2\delta' \eta' C X' \Omega^{1/2} \Gamma \Gamma' \Omega^{-1/2} (X \eta \delta + \xi) \\
&\quad - 2\sigma^2 \text{tr}(C X' \Omega^{1/2} \Gamma (\Gamma + \zeta \zeta') \Gamma' \Omega^{1/2} X S^{-1}) \\
&\quad + \sigma^2 \text{tr}(C X' \Omega^{1/2} \Gamma (\Gamma + \zeta \zeta') \Gamma' \Omega^{1/2} X C) \\
&= \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1}) + \xi' X S^{-2} X \xi + 2\delta' \eta' C X' \xi \\
&\quad + 2\delta' \eta' \eta \delta - 2\sigma^2 \text{tr}(C X' \Omega X S^{-1}) \\
&\quad - 2(\delta' \eta' X' + \xi') \Omega^{-1/2} \Gamma \Gamma' \Omega^{1/2} X S^{-1} C X' \Omega^{1/2} \Gamma \Gamma' \Omega^{-1/2} (X \eta \delta + \xi) \\
&\quad + \sigma^2 \text{tr}(C X' \Omega X C) \\
&\quad + (\delta' \eta' X' + \xi') \Omega^{-1/2} \Gamma \Gamma' \Omega^{1/2} X C C X' \Omega^{1/2} \Gamma \Gamma' \Omega^{-1/2} (X \eta \delta + \xi) \\
&= \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1} - 2C X' \Omega X S^{-1} + C X' \Omega X C) \\
&\quad + 2\delta' \eta' \eta \delta + \xi' X S^{-2} X \xi + 2\delta' \eta' C X' \xi \\
&\quad - 2\delta' \eta' X' X S^{-1} C X' X \eta \delta - 2\delta' \eta' X' X S^{-1} C X' \xi \\
&\quad - 2\xi' X S^{-1} C X' X \eta \delta - 2\xi' X S^{-1} C X' \xi + \delta' \eta' X' X C C X' X \eta \delta \\
&\quad + 2\delta' \eta' X' X C C X' \xi + \xi' X C C X' \xi \\
&= \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1} - 2C X' \Omega X S^{-1} + C X' \Omega X C) \\
&\quad + 2\delta' \eta' \eta \delta + \xi' X S^{-2} X \xi + 2\delta' \eta' X' \xi - 2\delta' \eta' \eta \delta - 2\delta' \eta' C X' \xi \\
&\quad - 2\xi' X S^{-1} \eta \delta - 2\xi' X S^{-1} C X' \xi + \delta' \eta' \eta \delta + 2\delta' \eta' C X' \xi + \xi' X C C X' \xi \\
&= \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1} - 2C X' \Omega X S^{-1} + C X' \Omega X C) + \delta' \eta' \eta \delta \\
&\quad + \xi' X (C - S^{-1}) X' \xi + 2\delta' \eta' (C - S^{-1}) X' \xi \\
&= \rho(b^*, \beta).
\end{aligned}$$

□

Proof of Corollary 5.3.

We require the following preliminary lemma.

Lemma 5A.1.

When $\Omega = I_T$

$$P_{mi} = h'(m-1, 0) \quad ; \quad i = 1, \dots, J.$$

Proof.

When $\Omega = I_T$

$$\begin{aligned} P_{mi} &= \Pr. \left(\lambda_i \chi_{(m; \theta_i)}^{2'} + \sum_{j=1}^T \lambda_j \chi_{(1; \theta_j)}^{2'} \leq 0 \right) \\ &= \Pr. \left(\lambda_i \chi_{(m; \theta_i)}^{2'} + \sum_{j=1}^J \lambda_j \chi_{(1; \theta_j)}^{2'} + \sum_{j=J+1}^K \lambda_j \chi_{(1; \theta_j)}^{2'} + \sum_{j=K+1}^T \lambda_j \chi_{(1; \theta_j)}^{2'} \leq 0 \right); \quad i=1, \dots, J \\ &= \Pr. \left(\sigma^2 \chi_{(m; \theta_i)}^{2'} + \sum_{j=1}^J \sigma^2 \chi_{(1; \theta_j)}^{2'} + \sum_{j=J+1}^K 0 \chi_{(1; \theta_j)}^{2'} - \sum_{j=K+1}^T \frac{cJ}{T-K} \sigma^2 \chi_{(1; \theta_j)}^{2'} \leq 0 \right); \quad i=1, \dots, J \\ &= \Pr. \left(\chi_{(J+m-1; \lambda_i)}^{2'} - \frac{cJ}{T-K} \chi_{(T-K; \lambda_2)}^{2'} \leq 0 \right) \quad i=1, \dots, J \\ &= h'(m-1, 0) \quad i=1, \dots, J \quad \square \end{aligned}$$

by Lemmas 4.1 and 4.2.

Proof of Corollary 5.3

a) $B(\hat{X}\beta, E(y)) = -M\xi - X C X' \Upsilon P_3 \Upsilon' (X\eta\delta + \xi)$

When $\Omega = I_T$ we have $\Upsilon = \bar{\Upsilon}$, the joint eigenvector matrix of $X C X'$, $X S^{-1} X'$ and $(X C X' - \frac{cJ}{T-K} M)$ by Lemma 4.1. Hence $\Upsilon' X C X' \Upsilon = \bar{I}_J$. Also, $P_{31} = h'(2, 0)$ for $i = 1, \dots, J$ by Lemma 5A.1.

Therefore,

$$\begin{aligned} B(\hat{X}\beta, E(y)) &= M\xi - \Upsilon \bar{\Upsilon}' X C X' \Upsilon P_3 \Upsilon' (X\eta\delta + \xi) \\ &= -M\xi - \Upsilon \bar{I}_J P_3 \Upsilon' (X\eta\delta + \xi) \end{aligned}$$

$$\begin{aligned}
&= -M\xi - \bar{Y}\bar{I}_J Y' (X\eta\delta + \xi)h'(2,0) \\
&= -M\xi - Y Y' X C X' Y Y' (X\eta\delta + \xi)h'(2,0) \\
&= -(M + X C X' h(2,0))\xi - X\eta\delta h'(2,0).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\rho(\hat{X}\hat{\beta}, E(y)) &= \sigma^2 \text{tr}(X' X S^{-1}) + \xi' M \xi - \sigma^2 \text{tr}(X C X' Y B Y') \\
&\quad + 2(X\eta\delta + \xi)' X C X' Y P_3 Y' (X\eta\delta + \xi).
\end{aligned}$$

Consider

$$\begin{aligned}
\text{tr}(X C X' Y B Y') &= \text{tr}(Y Y' X C X' Y Y' X C X' Y B Y') \\
&= \text{tr}(\bar{I}_J B \bar{I}_J) \\
&= \text{tr}(\bar{I}_J h'(2,0) + \sigma^{-2} Y' (\xi + X\eta\delta) (\delta' \eta' X' + \xi') Y \bar{I}_J) h'(4,0) \\
&= J h'(2,0) + (\delta' \eta' X' + \xi') Y Y' X C X' Y Y' (\xi + X\eta\delta) h'(4,0) \\
&= J h'(2,0) + 2\lambda_1 h'(4,0),
\end{aligned}$$

when $\Omega = I_T$.

Also,

$$\begin{aligned}
(X\eta\delta + \xi)' X C X' Y P_3 Y' (X\eta\delta + \xi) &= (X\eta\delta + \xi)' Y Y' X C X' Y P_3 Y' (X\eta\delta + \xi) \\
&= (X\eta\delta + \xi)' X C X' (X\eta\delta + \xi) h'(2,0) \\
&= 2\sigma^2 \lambda_1 h'(2,0)
\end{aligned}$$

hence

$$\rho(\hat{X}\hat{\beta}, E(y)) = \sigma^2 (K + 2\lambda_2 + (4\lambda_1 - J)h'(2,0) - 2\lambda_1 h'(4,0))$$

when $\Omega = I_T$.

$$\begin{aligned}
\text{b) } B(\hat{\beta}, \beta) &= S^{-1} X' \xi - C X' Y P_3 Y' (X\eta\delta + \xi) \text{ when } \Omega = I_T \\
&= S^{-1} X' \xi - S^{-1} X' Y Y' X C X' Y P_3 Y' (X\eta\delta + \xi) \\
&= S^{-1} X' \xi - S^{-1} X' Y \bar{I}_J Y' (X\eta\delta + \xi) h'(2,0) \\
&= S^{-1} X' \xi - C X' (X\eta\delta + \xi) h'(2,0)
\end{aligned}$$

$$= (S^{-1}X' - CX' h'(2,0)) \xi - \eta \delta h'(2,0), \text{ when } \Omega = I_T.$$

Similarly,

$$\begin{aligned} \rho(\hat{\beta}, \beta) &= \sigma^2 \text{tr}(S^{-1}X'XS^{-1}) + \xi XS^{-2}X'\xi + 2\delta' \eta' CX' \Upsilon P_3 \Upsilon' (X\eta\delta + \xi) \\ &\quad - 2\sigma^2 \text{tr}(CX' \Upsilon B \Upsilon' XS^{-1}) + \sigma^2 \text{tr}(CX' \Upsilon B \Upsilon' XC), \text{ when } \Omega = I_T. \end{aligned}$$

Now,

$$\begin{aligned} 2\delta' \eta' CX' \Upsilon P_3 \Upsilon' (X\eta\delta + \xi) &= 2\delta' \eta' S^{-1}X' \Upsilon \Upsilon' XCX' \Upsilon P_3 \Upsilon' (X\eta\delta + \xi) \\ &= 2\delta' \eta' CX' (X\eta\delta + \xi) h'(2,0) \\ &= 2(\delta' \eta' \eta \delta + \delta' \eta' CX' \xi) h'(2,0). \end{aligned}$$

Hence, when $\Omega = I_T$

$$\begin{aligned} \text{tr}(CX' \Upsilon B \Upsilon' XS^{-1}) &= \text{tr}(S^{-1}X' \Upsilon \Upsilon' XCX' \Upsilon B \Upsilon' XS^{-1}) \\ &= \text{tr}(S^{-1}X \Upsilon' \bar{I}_J B \Upsilon XS^{-1}) \\ &= \text{tr} \left(S^{-1}X \Upsilon' \bar{I}_J B \bar{I}_J \Upsilon X' S^{-1} \right) \\ &\quad + \sigma^{-2} \text{tr} \left(S^{-1}X \Upsilon' \bar{I}_J \xi \xi' \bar{I}_K \Upsilon X' S^{-1} \right) h'(0,0) \\ &\quad + \text{tr} \left(S^{-1}X \Upsilon' \bar{I}_J \xi \xi' \bar{I}_T \Upsilon X' S^{-1} \right) h'(0,2). \end{aligned}$$

Also,

$$\begin{aligned} \text{tr}(CX' \Upsilon B \Upsilon' XC) &= \text{tr} \left(S^{-1}X \Upsilon' \bar{I}_J B \bar{I}_J \Upsilon X' S^{-1} \right) \\ &= \text{tr}(C) h'(2,0) \\ &\quad + \sigma^{-2} \text{tr} \left(CX' (\xi + X\eta\delta) (\delta' \eta' X' + \xi') XC \right) h'(4,0) \\ &= \text{tr}(C) h'(2,0) \\ &\quad + \sigma^{-2} (\delta' + \xi' XS^{-1}R') \eta' \eta (RS^{-1}X' \xi + \delta) h'(4,0), \end{aligned}$$

when $\Omega = I_T$.

Therefore,

$$\rho(\hat{\beta}, \beta) = \sigma^2 \text{tr}(S^{-1}) + \xi' X S^{-2} X' \xi + 2\sigma^2 (\delta' \eta' \eta \delta + \delta' \eta' C S' \xi) h \quad (2,0)$$

$$- \sigma^2 \text{tr}(C) h' \quad (2,0)$$

$$- (\delta' + \xi' X S^{-1} R') \eta' \eta (R S^{-1} X' \xi + \delta) h' \quad (4,0)$$

$$- 2\xi' \bar{I}_K \Gamma X' S^{-2} X \Gamma' \bar{I}_J \xi h' \quad (0,0)$$

$$- 2\xi' \bar{I}_T \Gamma X' S^{-2} X \Gamma' \bar{I}_J \xi h' \quad (0,2)$$

□

CHAPTER SIX

PRE-TEST ESTIMATION OF THE SCALE PARAMETER IN A LINEAR MODEL WITH A MIS-SPECIFIED ERROR COVARIANCE MATRIX

6.1 Introduction

Although historically less attention has been focussed on the estimation of the regression scale parameter, σ^2 , after a pre-test, compared with the pre-test estimation of the coefficient vector or the pre-test predictor, in practical terms it is necessary to estimate σ^2 if further hypothesis tests are to be carried out or if an applied researcher is interested in an analysis of the precision of estimates obtained. Intuitively we would expect a mis-specification of the error covariance matrix to have a major effect on an estimator of the scale parameter as the estimator would be less precise in the presence of uncorrected autocorrelation or heteroscedasticity than would otherwise be the case. It is this problem which we consider here.

In section 6.2 we derive formulae describing the risk and bias of the unrestricted, restricted and pre-test estimators in the case of a general error covariance matrix. As these formulae are somewhat complicated, a special case in which the formulae may be simplified is examined in section 6.3. The formulae are evaluated numerically in this case, as described in section 6.4. Section 6.5 concludes the chapter.

6.2 Properties of the Component Estimators

Consider the linear regression model described by

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \Omega)$$

where ϵ is a $T \times 1$ random vector, Ω is a positive definite, symmetric $T \times T$ matrix, y and X are a $T \times 1$ vector and a $T \times K$ matrix of observations respectively, with X non-stochastic and of rank K , and β is a $K \times 1$ vector of unknown coefficients.

Assume that this model is estimated as $y = X\beta + \epsilon$ where ϵ is wrongly assumed to be $N(0, \sigma^2 I_T)$ distributed. J , non-stochastic linear restrictions, to be tested, are described by the hypotheses $H_0: R\beta = r$ vs $H_A: R\beta \neq r$, where R is a $J \times K$ matrix of rank J ($\leq K$). As usual, these restrictions are tested using the Wald statistic described by

$$u = \frac{(Rb-r)'(RS^{-1}R')^{-1}(Rb-r)(T-K)}{(y-Xb)'(y-Xb)J},$$

where $S = X'X$ and $b = S^{-1}X'y$. The properties of this statistic under these conditions are examined in Chapter 4.

If the test statistic, u , takes a value greater than the critical value, c , the null hypothesis is rejected and the scale parameter is estimated by the unrestricted estimator, s_1^2 , otherwise the restrictions are not rejected and the scale parameter is estimated by the restricted estimator, s_1^{*2} . This procedure gives rise to the pre-test estimator,

$$\hat{\sigma}_1^2 = \begin{cases} s_1^{*2} = (T+\Gamma)^{-1}(y-Xb^*)'(y-Xb^*) & ; \quad u < c \\ s_1^2 = (T+\Delta)^{-1}(y-Xb)'(y-Xb) & ; \quad u \geq c \end{cases},$$

where s_1^2 , s_1^{*2} and $\hat{\sigma}_1^2$ describe the maximum likelihood, ML, least squares, LS, and "minimum mean squared error", MS, estimators as $(\Delta, \Gamma) = (0, 0)$, $(-K, -\frac{1}{2}K)$ or $(2-K, 2-\frac{1}{2}K)$ respectively.¹ (That is, we set $i = \text{ML, LS, MS.}$)

¹ Note that the MS estimators will not generally be the actual minimum mean squared error estimators in a mis-specified model.

To derive the relative bias and relative risk functions of the unrestricted and restricted estimators, we require the following lemma.

Lemma 6.1.

If w is a $T \times 1$ random vector with $w \sim N(\mu, V)$ and A is a general $T \times T$ matrix, then

$$E[(w'Aw)]^2 = (\text{tr}(AV))^2 + 2\text{tr}(AV)^2 + 4\mu'AVA\mu + 2\mu'A\mu\text{tr}(AV) + (\mu'A\mu)^2$$

Proof.

See Appendix 6A.

Following the convention adopted in the literature, we will consider the relative bias and risk under quadratic loss of these estimators. These are defined as

$$B(\hat{\sigma}^2, \sigma^2) = \frac{1}{\sigma^2}(E(\hat{\sigma}^2) - \sigma^2)$$

and

$$\rho(\hat{\sigma}^2, \sigma^2) = \frac{1}{\sigma^4}[E(\hat{\sigma}^2 - \sigma^2)^2] = \frac{1}{\sigma^4}(E(\hat{\sigma}^2) - 2\sigma^2E(\hat{\sigma}^2) + \sigma^4)$$

respectively, where $\hat{\sigma}^2$ is an estimator of σ^2 . For brevity, we shall, on occasion, drop the qualifier "relative" when referring to bias and risk.

Theorem 6.1.

The relative bias and risk of the unrestricted and restricted estimators of the scale parameter are given by

$$\text{a) } B(s_1^2, \sigma^2) = \frac{1}{\sigma^2}(E(s_1^2) - \sigma^2) = (T+\Delta)^{-1}\text{tr}(M\Omega) - 1$$

$$\text{b) } \rho(s_1^2, \sigma^2) = \frac{1}{\sigma^4}[E(s_1^2 - \sigma^2)^2] = (T+\Delta)^{-2} \left[(\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 - 2(T+\Delta)\text{tr}(M\Omega) + (T+\Delta)^2 \right],$$

$$\text{c) } B(s_1^{*2}, \sigma^2) = (T+\Gamma)^{-1} \left(\text{tr}(\phi\Omega) + \frac{1}{\sigma^2} \delta' (RS^{-1}R')^{-1} \delta \right) - 1, \text{ and}$$

$$\begin{aligned}
d) \quad \rho(s_1^{*2}, \sigma^2) = (T+\Gamma)^{-2} & \left[\left(\text{tr}(\phi\Omega) \right)^2 + 2\text{tr}(\phi\Omega)^2 + \frac{1}{\sigma^2} \left(4\delta' \eta' X' \Omega X \eta \delta \right. \right. \\
& + 2\delta' (RS^{-1}R')^{-1} \delta \text{tr}(\phi\Omega) \left. \right) + \frac{1}{\sigma^4} \left(\delta' (RS^{-1}R')^{-1} \delta \right)^2 \\
& \left. - 2(T+\Gamma) \left(\text{tr}(\phi\Omega) + \frac{1}{\sigma^2} \delta' (RS^{-1}R')^{-1} \delta \right) + (T+\Gamma)^2 \right]
\end{aligned}$$

respectively, where $\phi = XCX' + M$, $\delta = R\beta - r$, $\eta = S^{-1}R'(RS^{-1}R')^{-1}$, $M = I_T - XS^{-1}X'$ and $C = S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$. Note that $\frac{1}{2\sigma^2}(\delta'(RS^{-1}R')^{-1}\delta) = \lambda$, the numerator non-centrality parameter of the distribution of the test statistic in the case where $\Omega = I_T$.

Proof.

See Appendix 6A.

It is apparent from the above formulae that the restricted estimator generally is more biased than the unrestricted estimator with both the bias and risk of the restricted estimator increasing without bound as the level of hypothesis error, δ , increases. In a well specified model the risk of the restricted estimator is always less than the unrestricted estimator's risk when the restrictions are valid, but it is not obvious from the risk functions that this condition will continue to hold in a model with a mis-specified error covariance matrix.

Corollary 6.1

When $\Omega = I_T$ the formulae for the bias and risk of the unrestricted and restricted estimator collapse to

- a) $B(s_1^2, \sigma^2) = -(T+\Delta)^{-1}(K+\Delta)$,
- b) $\rho(s_1^2, \sigma^2) = (T+\Delta)^{-2}(2\nu+(K+\Delta)^2)$,
- c) $B(s_1^{*2}, \sigma^2) = (T+\Gamma)^{-1}(J-K-\Gamma+2\lambda)$ and
- d) $\rho(s_1^{*2}, \sigma^2) = (T+\Gamma)^{-2}(2(J+\nu+4\lambda)+(J-K-\Gamma+2\lambda)^2)$

respectively, where² $\nu = T-K$.

Proof.

See Appendix 6A.

6.3 The Bias and Risk of the Pre-Test Estimators

Consider the pre-test estimator of the scale parameter. This estimator can be written as

$$\hat{\sigma}_1^2 = \begin{cases} (T+\Gamma)^{-1} \bar{\epsilon}' (XCX' + M) \bar{\epsilon} & ; u < c \\ (T+\Delta)^{-1} \bar{\epsilon}' M \bar{\epsilon} & ; u \geq c \end{cases}$$

where $\bar{\epsilon} = \epsilon + X\eta\delta$. Recall from Theorem 4.1 that in this case $u < c \Leftrightarrow z' \Lambda z < 0$, where $z \equiv \sigma^{-1} \Gamma' \Omega^{-1/2} (\epsilon + X\eta\delta)$, $z \sim N(\zeta, I_T)$, $\Lambda = \text{diag}\{\lambda_i\}$ and λ_i and Γ are the eigenvalues and eigenvectors of the matrix $\Omega^{1/2} (XCX' - \frac{Jc}{(T-K)} M) \Omega^{1/2}$. Therefore we can write the pre-test estimator as

$$\hat{\sigma}_1^2 = (T+\Delta)^{-1} \bar{\epsilon}' M \bar{\epsilon} + \left[(T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) \bar{\epsilon}' M \bar{\epsilon} + (T+\Gamma)^{-1} \bar{\epsilon}' XCX' \bar{\epsilon} \right] \Psi(z' \Lambda z),$$

where $\Psi(x)$ is an indicator function such that

$$\Psi(x) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } x \geq 0 \end{cases}.$$

If we define B as above in Theorem 4.2, a $T \times T$ matrix with ij 'th element given by

$$B_{ij} = \begin{cases} P_{31} + \zeta_1^2 P_{51} & ; \text{ when } i = j \\ \zeta_1 \zeta_j P_{31, 3j} & ; \text{ when } i \neq j \end{cases}$$

$$P_{mi} = \text{Pr.} \left(\lambda_i \chi_{(m; \theta_i)}^2 + \sum_{j \neq i}^T \lambda_j \chi_{(1; \theta_j)}^2 < 0 \right) \quad i = 3, 5$$

we have

² These formulae are equivalent to the ones given by Clarke *et al.* (1987a, 1987b) for the case of a well-specified model.

Theorem 6.2.

Under the stated assumptions the bias of the PTE of the scale parameter is given by

$$B(\hat{\sigma}_1^2, \sigma^2) = (T+\Delta)^{-1} \text{tr}(M\Omega) + (T+\Gamma)^{-1} \left((T+\Delta)^{-1} (\Delta-\Gamma) \text{tr}(M\Omega^{1/2} \Upsilon B \Upsilon' \Omega^{1/2}) \right. \\ \left. + \text{tr}(X C X' \Omega^{1/2} \Upsilon B \Upsilon' \Omega^{1/2}) \right) - 1.$$

Proof.

See Appendix 6A.

The derivation of the risk of the PTE is somewhat more complicated and we require the following preliminary lemma.

Lemma 6.2.

Given any two $T \times T$ symmetric matrices V and W

$$E[z' V z z' W z \Psi(z' \Lambda z)] = \sum_{i=1}^T \left[V_{ii} W_{ii} \omega_1^a + \sum_{j \neq i} \left[2(V_{ii} W_{ji} + V_{ji} W_{ii}) \omega_{ij}^b \right. \right. \\ \left. \left. + (V_{ii} W_{jj} + 2V_{ji} W_{ji}) \omega_{ij}^c \right. \right. \\ \left. \left. + \sum_{h \neq i, j} [(V_{ii} W_{jk} + 2V_{ji} W_{ki} + 2V_{ij} W_{ik} + V_{jk} W_{ii}) \omega_{ijk}^d \right. \right. \\ \left. \left. + \sum_{\ell \neq i, j, k} V_{ji} W_{\ell k} \omega_{ijk\ell}^e] \right] \right]$$

where,

$$\omega_1^a = 3P_{51} + 6\zeta_1^2 P_{71} + \zeta_1^4 P_{91},$$

$$\omega_{ij}^b = 3\zeta_i \zeta_j P_{51,3j} + \zeta_i^3 \zeta_j P_{71,3j} \quad ; i \neq j,$$

$$\omega_{ij}^c = P_{31,3j} + \zeta_i^2 P_{51,3j} + \zeta_j^2 P_{31,5j} + \zeta_i^2 \zeta_j^2 P_{51,5j} \quad ; i \neq j,$$

$$\omega_{ijk}^d = \zeta_j \zeta_k P_{31,3j,3k} + \zeta_i^2 \zeta_j \zeta_k P_{51,3j,3k} \quad ; i \neq j, k, j \neq k,$$

$$\omega_{ijk\ell}^e = \zeta_i \zeta_j \zeta_k \zeta_\ell P_{31,3j,3k,3\ell} \quad ; i \neq j, k, \ell, j \neq k, k \neq \ell,$$

with $P_{m,n,j,ok,p\ell} = \Pr. \left(\lambda_1 \chi_{(m;\theta_1)}^{2'} + \lambda_j \chi_{(m;\theta_j)}^{2'} + \lambda_k \chi_{(o;\theta_k)}^{2'} + \lambda_\ell \chi_{(p;\theta_\ell)}^{2'} \right.$
 $\left. + \sum_{h \neq 1, j, k, \ell}^T \lambda_k \chi_{(1;\theta_h)}^{2'} < 0 \right),$

and i, j, k and ℓ all distinct. In order to simplify the notation to some extent, if any of m, n, o or p is equal to unity it is not shown in the subscript; i.e., $P_{31} = P_{31,1j,1k,1\ell}$ etc.

Proof.

See appendix 6A.

Now, let us define

$$\Xi_1 \equiv E \left(z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z \Psi(z' \Lambda z) \right)$$

$$\Xi_2 \equiv E \left(z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z z' \Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon z \Psi(z' \Lambda z) \right)$$

and

$$\Xi_3 \equiv E \left(z' \Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon z z' \Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon z \Psi(z' \Lambda z) \right).$$

These can be evaluated, for a given data set, using Lemma 6.2.

Theorem 6.3.

Under the stated assumptions the risk of the PTE of the scale parameter is given by

$$\rho(\hat{\sigma}_1^2, \sigma^2) = ((T+\Delta)(T+\Gamma))^{-2} \left\{ (T+\Gamma)^2 \left((\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 \right) + (\Delta+\Gamma)(2T+\Delta+\Gamma)\Xi_1 + 2(T+\Delta)^2\Xi_2 \right.$$

$$\left. + (T+\Delta)^2\Xi_3 \right\} - 2 \left\{ (T+\Delta)^{-1} \text{tr}(M\Omega) + (T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) \text{tr}(M\Omega^{1/2} \Upsilon B \Upsilon' \Omega^{1/2}) \right.$$

$$\left. + (T+\Gamma)^{-1} \text{tr}(X C X' \Omega^{1/2} \Upsilon B \Upsilon' \Omega^{1/2}) \right\} + 1.$$

Proof.

See Appendix 6A.

The bias and risk of the PTE depend on the data, the form and validity of the restrictions, and the choice of the critical value through the probability values in Ξ_1 , Ξ_2 , Ξ_3 and B. Because these expressions are complicated, it is difficult to determine the effect of the mis-specification on the properties of the PTE. Some points can, however be made.

Corollary 6.2.

As the critical value chosen for the pre-test, c , tends towards zero (∞), the bias and risk of the PTE tend towards the bias and risk of the unrestricted (restricted) estimator.

Proof.

See Appendix 6A.

Intuitively this is because the probability of rejecting the null hypothesis tends to unity (zero) as the critical value tends towards zero (∞) and hence the unrestricted (restricted) estimator will always be used. In general, however, the properties of the test statistic will differ from the properties of its two component estimators.

Corollary 6.3.

When $\Omega = I_T$ the bias and risk of the PTE are given by³

$$B(\hat{\sigma}_1^2, \sigma^2) = (T+\Gamma)^{-1}(T+\Delta)^{-1} \left[(T+\Delta) (2\lambda h(4,0) + Jh(2,0)) \right. \\ \left. - (T+\Gamma)(K+\Delta) + \nu(\Delta-\Gamma)h(0,2) \right] \\ \rho(\hat{\sigma}_1^2, \sigma^2) = 1 + ((T+\Gamma)(T+\Delta))^{-2} \left(4\lambda(T+\Delta)^2 (\lambda h(8,0) + (J+2)h(6,0) + \nu h(4,0)) \right. \\ \left. - (T+\Gamma)h(4,0) + \nu(\nu+2)(T+\Gamma)^2 - 2(T+\Gamma)(T+\Delta) (\nu(T+\Gamma)) \right)$$

³ These expressions are equivalent to those given by Clarke *et al.* (1987a, 1987b) for the case of the correctly specified model.

$$+ \nu(\Delta-\Gamma)h(0,2)+J(T+\Delta)h(2,0)+J(T+\Delta)^2(2\nu h(2,2)+(J+2)h(4,0)) \\ + \nu(\nu+2)(\Delta-\Gamma)(2T+\Delta+\Gamma)h(0,4)\Bigg].$$

where

$$h(i,j) = \Pr. \left(\frac{\chi_{(J+1;\lambda)}^2}{\chi_{(T-K+J)}^2} < \frac{cJ}{T-K} \right)$$

Proof.

See Appendix 6A.

Given the complexity of the formulae for the pre-test bias and risk, it is difficult to make further observations regarding them without making numerical evaluations. Note, however, that in general this involves calculating between T^3 and T^4 different probability values for each possible value of θ , and each data generating process. Hence the amount of computer time required to evaluate the formulae is a significant constraint on the number of cases we can evaluate. In the next section we will consider a special case in which the formulae simplify somewhat, thus easing the computational burden.

6.4 Numerical Evaluation

6.4.1 The Model

Consider the model

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \equiv X\beta + \epsilon$$

where y and y_1 are $T \times 1$ and $T_1 \times 1$ vectors of observations on the dependent variable, X and X_1 are $T \times K$ and $T_1 \times K_1$, non-stochastic, matrices of observations on the explanatory variables, of rank K and K_1 respectively, β and β_1 are $K \times 1$ and $K_1 \times 1$ vectors of unknown coefficients and ϵ and ϵ_1 are $T \times 1$

and $T_1 \times 1$ disturbance vectors. In this case we will assume, for simplicity of exposition, that $T_1 = T_2 = \frac{1}{2}T$, $K_1 = K_2 = \frac{1}{2}K$ and

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N \left(0, \sigma^2 \begin{bmatrix} I_{T_1} & 0 \\ 0 & \psi I_{T_2} \end{bmatrix} \right).$$

The parameter ψ measures the degree of heteroscedasticity in the model. Consider a situation in which it is incorrectly assumed that $\varepsilon \sim N(0, \sigma^2 I_T)$ and a Chow test for structural change is applied to the model prior to estimating the scale parameter, σ^2 . The null and alternative hypotheses are given by

$$H_0: R\beta = r \Leftrightarrow \beta_1 = \beta_2 \quad \text{vs} \quad H_A: R\beta \neq r \Leftrightarrow \beta_1 \neq \beta_2$$

with $R = [I_J; -I_J]$, $r = 0$, $J = \frac{1}{2}K$.

Lemma 6.3.

Under the assumptions of the model there exists a $T \times T$ orthonormal matrix denoted $\tilde{\Upsilon}$ which is the joint eigenvector matrix of $\Omega^{1/2}XCX'\Omega^{1/2}$, $\Omega^{1/2}M\Omega^{1/2}$ and $\Phi = \Omega^{1/2}(XCX' - \frac{cJ}{T-K}M)\Omega^{1/2}$. Let the i 'th column of $\tilde{\Upsilon}$ be denoted \tilde{v}_i .

Therefore

$$\tilde{\Upsilon}\Omega^{1/2}M\Omega^{1/2}\tilde{\Upsilon} = \Lambda^M$$

and

$$\tilde{\Upsilon}\Omega^{1/2}XCX'\Omega^{1/2}\tilde{\Upsilon} = \Lambda^C$$

where $\Lambda^M \equiv \text{diag}\{\lambda_i^M\}$ and $\Lambda^C \equiv \text{diag}\{\lambda_i^C\}$, with λ_i^M and λ_i^C the i 'th eigenvalues of $\Omega^{1/2}M\Omega^{1/2}$ and $\Omega^{1/2}XCX'\Omega^{1/2}$ respectively, and

$$\lambda_i^M = 0 \quad ; \quad i = 1, \dots, K,$$

$$\lambda_i^M = 1 \quad ; \quad i = K+1, \dots, \frac{1}{2}T+K,$$

$$\lambda_i^M = \psi \quad ; \quad i = \frac{1}{2}(T+K)+1, \dots, T,$$

$$\lambda_i^C \neq 0 \quad ; \quad i = 1, \dots, J$$

and

$$\lambda_1^C = 0 \quad ; \quad i = J+1, \dots, T.$$

Proof.

See Appendix 6A.

Using Lemma 6.3 it is straightforward to simplify the bias and risk functions of the PTE of the scale parameter.

Theorem 6.4.

In this case the bias and risk of the PTE of the scale parameter are given by

$$\begin{aligned} B(\hat{\sigma}_1^2, \sigma^2) = & (T+\Delta)^{-1} \left[\text{tr}(M\Omega) + (T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) \left(\sum_{i=K+1}^{(T+K)/2} P_{3i} + \psi \sum_{i=(T+K)/2}^T P_{3i} \right) \right. \\ & \left. + (T+\Gamma)^{-1} \sum_{i=1}^J \lambda_1^C (P_{3i} + \zeta_1^2 P_{5i}) \right] - 1 \end{aligned}$$

and

$$\begin{aligned} \rho(\hat{\sigma}_1^2, \sigma^2) = & ((T+\Delta)(T+\Gamma))^2 \left[(T+\Gamma)^2 \left((\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 \right) \right. \\ & + (2(T+\Gamma)(\Delta-\Gamma) + (\Delta-\Gamma)^2) \left(\sum_{i=K+1}^T (\lambda_i^M)^2 3P_{5i} + \sum_{\substack{j=1 \\ j \neq i}}^T \lambda_i^M \lambda_j^M P_{3i,3j} \right) \\ & + 2((T+\Gamma)(T+\Delta) + (\Delta-\Gamma)(T+\Delta)) \left(\sum_{i=1}^J \sum_{j=K+1}^T \lambda_i^C \lambda_j^M (P_{3i,3j} + \xi_i^2 P_{5i,3j}) \right) \\ & + (T+\Delta)^2 \left(\sum_{i=1}^J (\lambda_i^C)^2 (3P_{5i}) + 6\zeta_1^2 P_{7i} + \zeta_1^4 P_{9i} \right) + \sum_{\substack{j=1 \\ j \neq i}}^T \lambda_i^C \lambda_j^C (P_{3i,3j} + \zeta_1^2 P_{5i,3j}) \\ & + \zeta_j^2 P_{3i,5j} + \zeta_i^2 \zeta_j^2 P_{5i,5j} \Big) - 2 \left[(T+\Delta)^{-1} \text{tr}(M\Omega) + (T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) \sum_{i=K+1}^T \lambda_i^M P_{3i} \right. \\ & \left. + (T+\Gamma)^{-1} \sum_{i=1}^J \lambda_1^C (P_{3i} + \zeta_1^2 P_{5i}) \right] + 1 \end{aligned}$$

respectively.

Proof.

The proof follows from Theorem 6.3 and Lemma 6.3.

Note that these formulae, though apparently complex, are less computationally burdensome than those given in the previous section by two orders of magnitude. These formulae have been evaluated using a FORTRAN program written by the author incorporating Davies' (1980) algorithm and executed on a Vax 6340 computer. The data used are described in Appendix 4B.

6.4.2 Numerical Results

Typical unrestricted, restricted and PT estimator functions in a well-specified regression model are shown in Figures 6.1, 6.4 and 6.6. As noted in Chapter Two, in the case of the LS and MS estimators, there is a family of LS PTE's and a family of MS PTE's, which strictly dominate the corresponding unrestricted estimators. The minimum risk members of these families of LS and MS PT estimators are those associated with critical values of $c = 1$ and $c = \nu/(\nu+2)$ respectively. In particular, when $J \leq 2$ the LS PTE strictly dominates both of its component estimators. Quantitatively the risk of each of the estimators increases as ψ increases above unity. As ψ decreases below unity the estimator risks may either increase, or decrease, depending on the estimator being considered. The restricted estimator's risk generally decreases, as ψ decreases, while the unrestricted estimator's risk generally increases. The effect on the PTE depends, as we would expect, on the critical value of the pre-test. If the critical value is relatively low, the PTE tends to lie close to the unrestricted estimator and its risk increases as ψ decreases, and the converse occurs if the critical value is relatively large. Qualitatively the effect of the mis-specification is to increase the range of θ over which the restricted estimators have lower risk

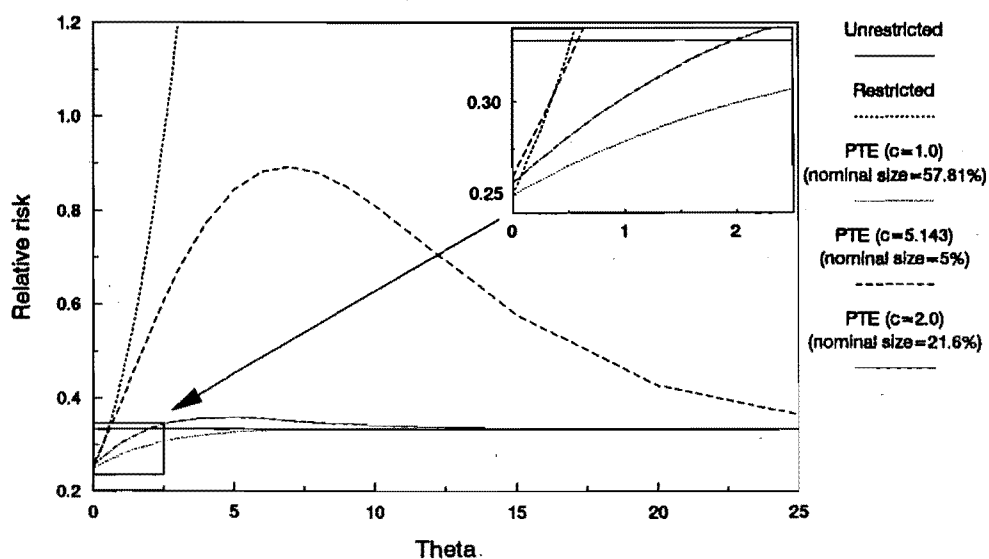
than the unrestricted estimators as ψ decreases below unity. This occurs regardless of which family of estimators is considered. The converse occurs as ψ increases above unity.

The effect of the mis-specification on the PTE depends on which of the LS, ML or MS families of estimators is being considered. As discussed in Chapter Four, one effect of the mis-specification is to introduce a distortion in the pre-test size and (size uncorrected) power. Intuitively we would expect such a distortion to affect the dominance of the PTE over the unrestricted estimator in the case of the LS and MS estimators and this is, in fact, the effect we observe.

Figure 6.1 illustrates the LS estimators in a well-specified model. The PTE associated with $c = 1$ is the dominating estimator, in this case as $J = 2$. Figure 6.2 illustrates the same model but in a case where there is a mis-specification of the error covariance matrix, in that there is uncorrected heteroscedasticity present.

FIGURE 6.1: Scale Parameter Estimator Risk Fns.

LS Estimators, Well Specified Model
Two regressors

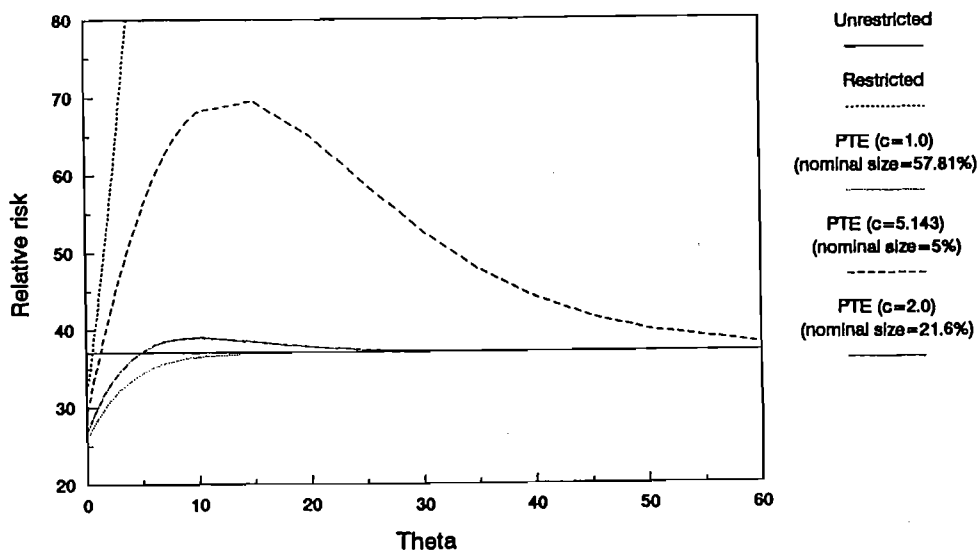


Application of Chow test

Sample size = 10

FIGURE 6.2: Scale Parameter Estimator Risk Fns.

LS Estimators, $\Psi = 10.0$
Regressors; constant and Australian M3.



Application of Chow test

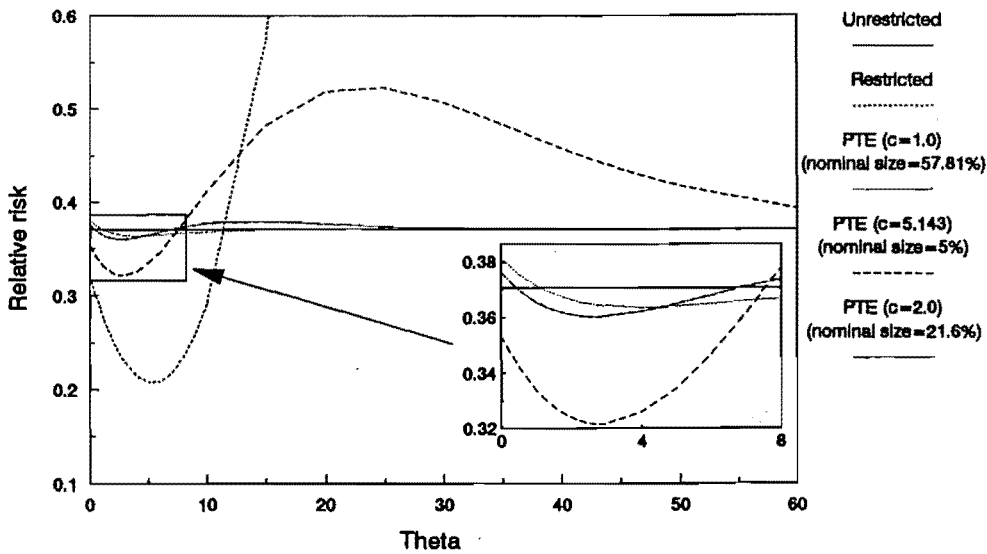
Sample size = 10

Regardless of the particular data used, the effect of increasing ψ above unity is qualitatively minor. The PTE associated with a critical value of $c = 1$ remains the dominating estimator and may become more attractive relative to the restricted estimator as the difference between the restricted and PT risks under the null hypothesis increases as ψ increases above unity.

Conversely, as ψ decreases below unity, the PTE associated with a critical value of $c = 1$ may no longer be the dominating estimator. This is true regardless of the particular form of the regressor data. As Figure 6.3 illustrates, not only is the PTE associated with $c = 1$ no longer the dominating estimator, but there is no choice of critical value which will give rise to a PTE that dominates the unrestricted estimator. This can be seen by considering the PT risk associated with $c = 2$.

FIGURE 6.3: Scale Parameter Estimator Risk Fns.

LS Estimators, $\Psi = 0.1$
Regressors; constant and Australian M3.



Application of Chow test

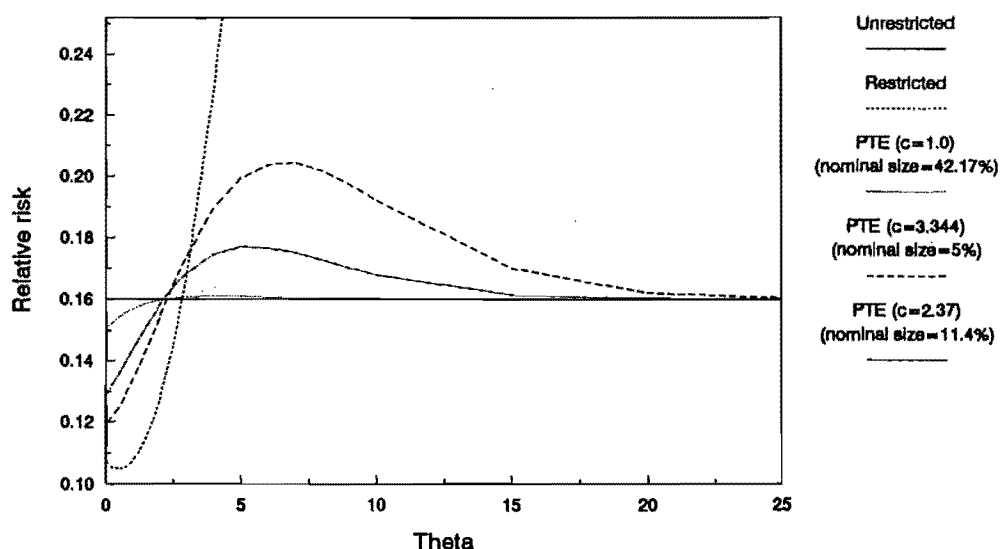
Sample size = 10

The PTE has a higher risk than the unrestricted estimator if the restrictions are valid, i.e., if $\theta = 0$. Any increase in c will reduce the PT risk at $\theta = 0$ as the PT risk will tend towards the restricted estimator risk, which is less than the unrestricted risk. However, the PTE associated with $c = 2$ has a higher risk than the unrestricted estimator at $\theta = 15$. Hence an increase in c will increase the risk of the PTE at this value of θ , therefore no dominating PTE exists.

The situation is reversed in the case of the ML estimators, regardless of the particular data under consideration. As Figure 6.4 shows, there is no PTE that strictly dominates either of its component estimators in the well-specified regression mode, also the PTE is never the minimum risk estimator. It is, in fact, the maximum risk estimator over some part of the parameter space. The risk functions of the ML estimators of the scale parameter display similar characteristics to those of the estimators of the coefficient vector.

FIGURE 6.4: Scale Parameter Estimator Risk Fns.

ML Estimators, Well Specified Model
Three regressors



Application of Chow test

Sample size = 20

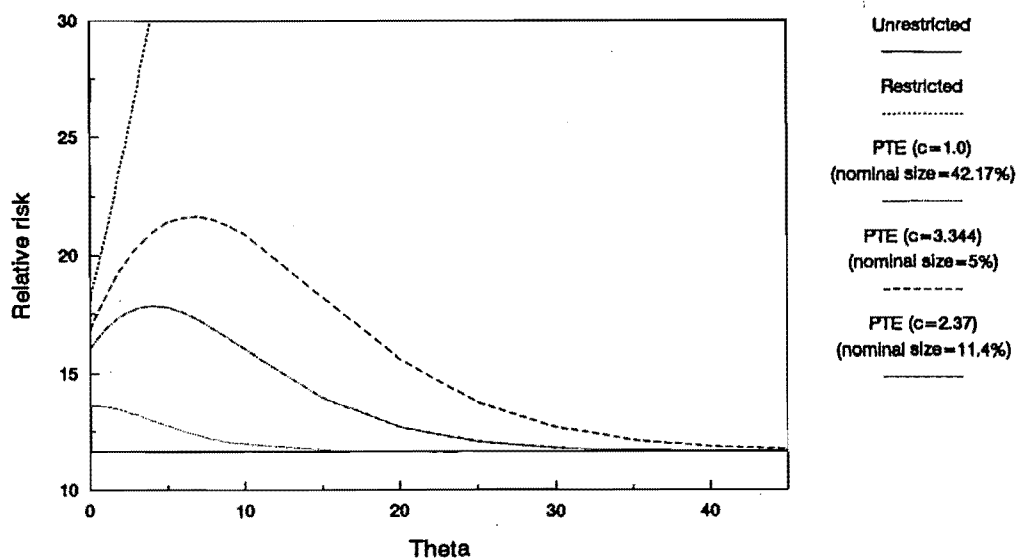
Figure 6.5 illustrates the effect on estimator risk of increasing ψ above unity. In contrast to the effect of uncorrected heteroscedasticity on the LS estimators, the ML, PT and restricted estimators may become strictly dominated by the unrestricted estimator in this case.⁴ Conversely, as ψ decreases below unity, the PTE may strictly dominate the unrestricted, but not the restricted, estimator, depending on the critical value of the pre-test and the degree of mis-specification. This situation is shown in Figure 6.6. In this instance the restricted estimator dominates the unrestricted estimator over a greater part of the parameter space than is assumed to be the case. Note also that PTE's associated with quite low nominal size may dominate the unrestricted estimator.

4

A similar result is found by Giles and Clarke (1989) in the case of applying the ML estimators of the scale parameter in a model with excluded relevant regressors. Mittelhammer (1984), Giles and Giles (1991) and Chapter Four of this thesis provide other examples of PTE's which are dominated by their unrestricted counterparts under the null hypothesis.

FIGURE 6.5: Scale Parameter Estimator Risk Fns.

ML Estimators, $\Psi = 10.0$
 Regressors; constant, trend and New Zealand GDP.

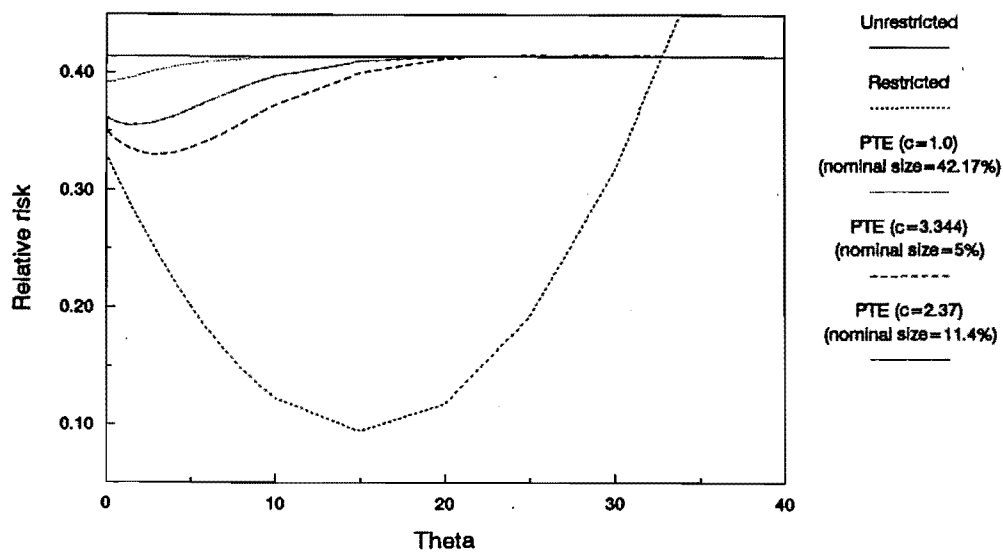


Application of Chow test

Sample size = 20

FIGURE 6.6: Scale Parameter Estimator Risk Fns.

ML Estimators, $\Psi = 0.1$
 Regressors; constant, trend and New Zealand GDP.

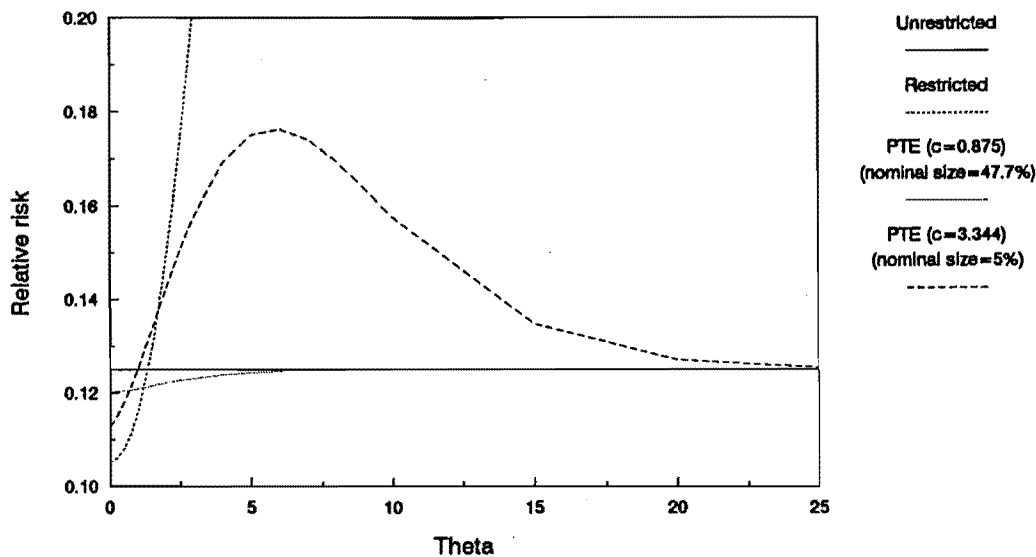


Application of Chow test

Sample size = 20

The effect of the mis-specification on the MS estimators is qualitatively the same as the effect on the LS estimators as Figures 6.7, 6.8 and 6.9 illustrate. Note that the 'optimal' critical value, in terms of minimizing the PTE risk, generally increases as ψ decreases below unity, although it is by no means as great as the critical value associated with traditional nominal test sizes of 5% or 1%.

FIGURE 6.7: Scale Parameter Estimator Risk Fns.
MS Estimators, Well Specified Model
Three regressors

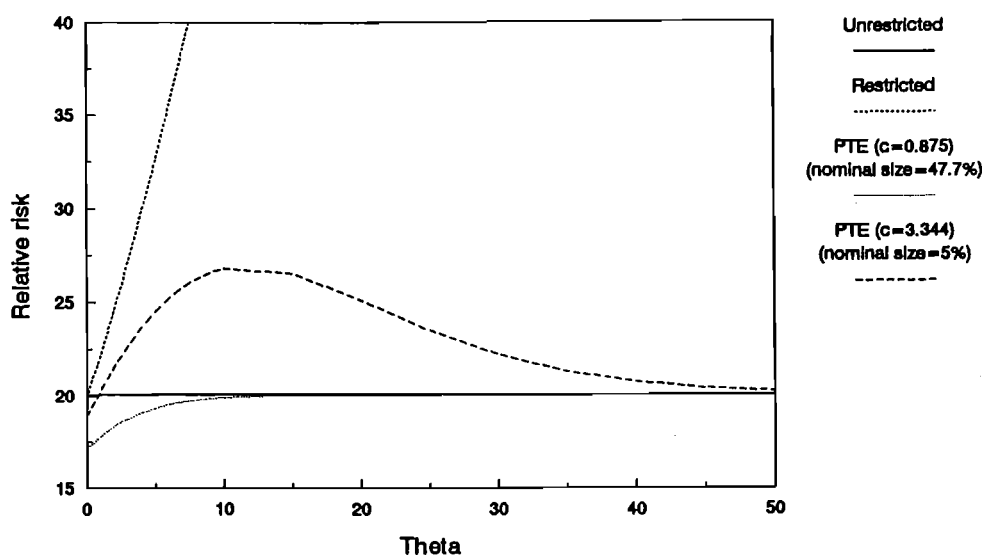


Application of Chow test

Sample size = 20

FIGURE 6.8: Scale Parameter Estimator Risk Fns.

MS Estimators, $\Psi = 10.0$
 Regressors; constant, trend and Aus\$-US\$ x-rate.

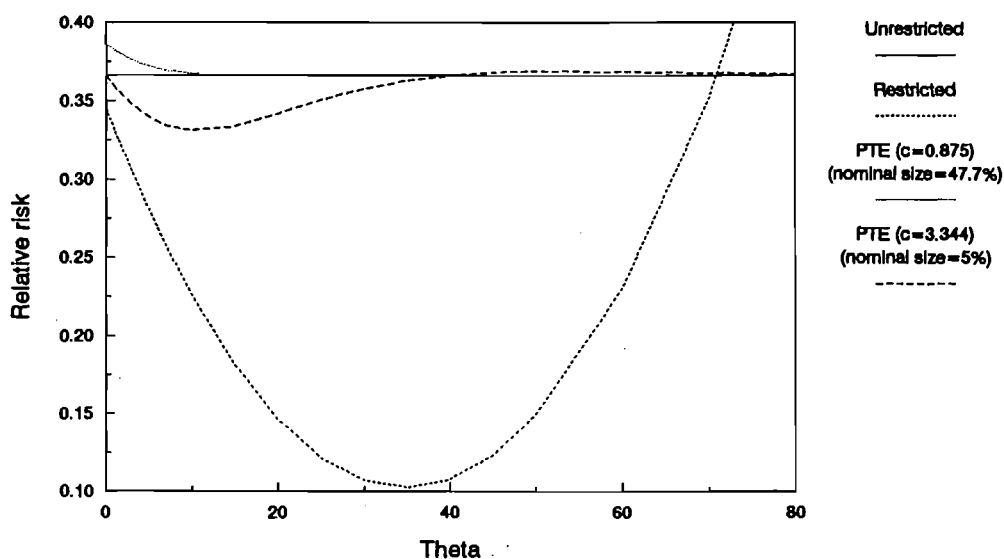


Application of Chow test

Sample size = 20

FIGURE 6.9: Scale Parameter Estimator Risk Fns.

MS Estimators, $\Psi = 0.01$
 Regressors; constant, trend and Aus\$-US\$ x-rate.



Application of Chow test

Sample size = 20

6.5 Conclusion

The practical implications of the heteroscedasticity mis-specification vary depending on the type of estimators we are considering and the degree of heteroscedasticity. However, it is apparent that if the LS or MS estimators are being used for σ^2 , with a critical value of unity or slightly greater, pre-testing is likely to remain the best strategy in terms of minimizing the risk, even in the presence of uncorrected heteroscedasticity. If, however, the ML estimators are applied, the PTE may well be strictly dominated.

Note, however, that the effect of the mis-specification on the MS estimators is quantitatively greater than the effect on the LS estimators, which is quantitatively greater than the effect on the ML estimators. That is, the ML estimators are the most robust of the three to the uncorrected heteroscedasticity. Thus, while the ML pre-test estimator is dominated by the ML unrestricted estimator, it strictly dominates both the MS and the LS family of estimators if the error covariance matrix is mis-specified in this way. On the basis of this study, it appears that the unrestricted ML estimator is the best estimator of those considered here if there is a concern that the errors may not be homoscedastic.

Appendix 6A

Proof of Lemma 6.1.

If w is a $T \times 1$ random vector with $w \sim N(\mu, V)$ and A is a general $T \times T$ matrix

$$E(w'Aw) = \text{tr}(AV) + \mu' A \mu \quad 6A.1$$

and

$$\text{covar}(w'Aw) = 2\text{tr}(AV)^2 + 4\mu' AVA\mu \quad 6A.2$$

from Searle (1982, p.351, 357).

The proof follows from noting that

$$\begin{aligned} \text{covar}(w'Aw) &= E \left[(w'Aw - E(w'Aw))^2 \right] \\ &= E \left[(w'Aw)^2 \right] - \left(E(w'Aw) \right)^2 \end{aligned} \quad \square$$

Proof of Theorem 6.1.

$$a) \quad s_1^2 = \frac{1}{(T+\Delta)} \varepsilon' M \varepsilon = \frac{1}{(T+\Delta)} \bar{\varepsilon}' M \bar{\varepsilon}$$

where $\bar{\varepsilon} = \varepsilon + X\eta\delta$, $\bar{\varepsilon} \sim N(X\eta\delta, \sigma^2\Omega)$

$$B(s_1^2, \sigma^2) = \frac{1}{\sigma^2} (E(s_1^2) - \sigma^2) = (T+\Delta)^{-1} \text{tr}(M\Omega) - 1$$

from 6A.1

$$\begin{aligned} b) \quad \rho(s_1^2, \sigma^2) &= \frac{1}{\sigma^4} \left(E[(s_1^2)^2] - 2E(s_1^2)\sigma^2 + \sigma^4 \right) \\ &= (T+\Delta)^{-2} \left((\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 \right) - 2(T+\Delta)^{-1} \text{tr}(M\Omega) + 1 \end{aligned}$$

by Lemma 6.1 and 6A.1.

c) We know that

$$\begin{aligned} \bar{\varepsilon}' X C X' \bar{\varepsilon} &= (Rb-r)' (RS^{-1}R')^{-1} (Rb-r) \\ &= (y-Xb^*)' (y-Xb^*) - (y-Xb)' (y-Xb), \end{aligned}$$

hence

$$(y-Xb^*)' (y-Xb^*) = \bar{\varepsilon}' X C X' \bar{\varepsilon} + \bar{\varepsilon}' M \bar{\varepsilon} = \bar{\varepsilon}' \phi \bar{\varepsilon},$$

where $\phi = (XCX' + M)$. Note that $\phi^2 = \phi$ and $\text{tr}(\phi) = J + T - K$.

$$B(s_1^{*2}, \sigma^2) = (T + \Gamma)^{-1} \left(\text{tr}(\phi\Omega) + \frac{1}{\sigma^2} \delta' \eta' X' (XCX' + M) X \eta \delta \right) - 1$$

by 6A.1, therefore

Hence

$$B(s_1^{*2}, \sigma^2) = (T + \Gamma)^{-1} \left(\text{tr}(\phi\Omega) + \frac{1}{\sigma^2} \delta' (RS^{-1}R')^{-1} \delta \right) - 1$$

$$\begin{aligned} \text{d) } E[(s_1^{*2}, \sigma^2)] &= (T + \Gamma)^{-2} \left(\sigma^4 (\text{tr}(\phi\Omega))^2 + 2\sigma^4 \text{tr}(\phi\Omega)^2 + \sigma^2 4\delta' \eta' X' \phi \Omega \phi X \eta \delta \right. \\ &\quad \left. + 2\sigma^2 \delta' \eta' X' \phi X \eta \delta + (\phi\Omega) + (\delta' \eta' X' \phi X \eta \delta)^2 \right) \end{aligned}$$

by Lemma 6.1.

Hence

$$\begin{aligned} \rho(s_1^{*2}, \sigma^2) &= (T + \Gamma)^{-2} \left[(\text{tr}(\phi\Omega))^2 + 2\text{tr}(\phi\Omega)^2 + \frac{4}{\sigma^2} \delta' \eta' X' \Omega X \eta \delta \right. \\ &\quad \left. + \frac{2}{\sigma^2} \delta' (RS^{-1}R')^{-1} \delta \text{tr} \phi \Omega + \frac{1}{\sigma^4} (\delta' (RS^{-1}R')^{-1} \delta)^2 \right] \\ &\quad - 2(T + \Gamma)^{-1} \left(\text{tr}(\phi\Omega) + \frac{1}{2\sigma^2} \delta' (RS^{-1}R')^{-1} \delta \right) + 1. \end{aligned}$$

□

Proof of Corollary 6.1.

When $\Omega = I_T$

$$\text{a) } B(s_1^2, \sigma^2) = (T + \Delta)^{-1} \text{tr}(M) - 1 = (T + \Delta)^{-1} (T - K(T + \Delta))$$

$$\begin{aligned} \text{b) } \rho(s_1^2, \sigma^2) &= (T + \Delta)^{-2} \left((\text{tr}(M))^2 + 2\text{tr}(M) - 2(T + \Delta)\text{tr}(M) + (T + \Delta)^{-2} \right) \\ &= (T + \Delta)^{-2} (T^2 - 2TK + K^2 - 2K - 2T^2 + 2TK - 2\Delta T + 2\Delta K + T^2 + 2\Delta T + \Delta^2) \\ &= (T + \Delta)^{-2} (2v + (K + \Delta)^2) \end{aligned}$$

$$c) \quad B(s_1^{*2}, \sigma^2) = (T+\Gamma)^{-1} (\text{tr}(XCX' + M)2\lambda) - 1$$

$$= (T+\Gamma)^{-1} (J+T-K+2\lambda-(T+\Gamma))$$

$$d) \quad \rho(s_1^{*2}, \sigma^2) = (T+\Gamma)^{-2} \left[(\text{tr}(\phi))^2 + 2\text{tr}(\phi) + \frac{4}{\sigma^2} \delta' (RS^{-1}R') RS^{-1}X' XS^{-1}R' (RS^{-1}R')^{-1} \delta \right.$$

$$\left. + \frac{2}{\sigma^2} \delta' (RS^{-1}R')^{-1} \delta \text{tr} \phi + \frac{1}{\sigma^4} (\delta' (RS^{-1}R')^{-1} \delta)^2 - 2(T+\Gamma) (\text{tr}(\phi) + 2\lambda) + (T+\Gamma)^2 \right]$$

$$= (T+\Gamma)^{-2} [(J+T-K) + 8\lambda + 4\lambda(J+T-K) + 4\lambda^2 - 2(T+\Gamma)(J+T-K) - 4(T+\Gamma)\lambda + (T+\Gamma)^2]$$

$$= (T+\Gamma)^{-2} [J^2 + 2JT - 2JK + T^2 - 2TK + K^2 + 2J + 2T - 2K + 8\lambda + 4\lambda J + 4\lambda T$$

$$- 4\lambda K + 4\lambda^2 - 2JT - 2T^2 + 2TK - 2\Gamma J - 2\Gamma T + 2\Gamma K - 4\lambda T + T^2 + 2\Gamma T + \Gamma^2]$$

$$= (T+\Gamma)^{-2} [2(J+T-K+4\lambda) + (J-K-\Gamma+2\lambda)^2] .$$

Proof of Theorem 6.2.

$$\hat{\sigma}_1^2 = (T+\Delta)^{-1} \bar{\epsilon}' M \bar{\epsilon} + \sigma^2 ((T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z$$

$$+ (T+\Gamma)^{-1} z' \Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon z) \Psi(z' \Lambda z),$$

hence

$$E(\hat{\sigma}_1^2, \sigma^2) = (T+\Delta)^{-1} E(\bar{\epsilon}' M \bar{\epsilon})$$

$$+ \sigma^2 \left((T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) \text{tr}(M \Omega^{1/2} \Upsilon E(z z' \Psi(z' \Lambda z) \Upsilon' \Omega^{1/2})) \right)$$

$$+ (T+\Gamma)^{-1} \text{tr}(X C X' \Omega^{1/2} \Upsilon E(z z' \Psi(z' \Lambda z) \Upsilon' \Omega^{1/2}))$$

$$= \sigma^2 (T+\Delta)^{-1} \text{tr}(M \Omega) + \sigma^2 ((T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) \text{tr}(M \Omega^{1/2} \Upsilon B \Upsilon' \Omega^{1/2})$$

$$+ (T+\Gamma)^{-1} \text{tr}(X C X' \Omega^{1/2} \Upsilon B \Upsilon' \Omega^{1/2}))$$

□

by Lemma 4A.1.

Proof of Lemma 6.2.

To prove this lemma we will first prove the following result. If a random variable q is $N(\mu, 1)$ distributed and $f(q^2)$ is a single valued real function then

$$E(q^3 f(q^2)) = E[3\mu f(\chi_{(5;\mu^2/2)}^{2'}) + \mu^3 f(\chi_{(7;\mu^2/2)}^{2'})] \quad 6A.1$$

Proof.

$h(q^2) = q^2 f(q^2)$ is a single valued real function therefore

$$E(qh(q^2)) = \mu E[2\mu f(\chi_{(3;\mu^2/2)}^{2'})]$$

by Lemma 2 of Judge and Bock (1978, p.320).

Hence

$$\begin{aligned} E(q^3 f(q^2)) &= \mu E[h(\chi_{(3;\mu^2/2)}^{2'})] = \mu E[\chi_{(3;\mu^2/2)}^{2'} f(\chi_{(3;\mu^2/2)}^{2'})] \\ &= \mu E[3f(\chi_{(5;\mu^2/2)}^{2'}) + 2(\frac{\mu^2}{2})f(\chi_{(7;\mu^2/2)}^{2'})] \end{aligned}$$

by Lemma 1 of Clarke *et al.* (1987a). □

Using this result we can show the following:

$$a) \quad E(z_1^4 \Psi(z' \Lambda z)) = \omega_1^a \quad 6A.2$$

$$b) \quad E(z_1^3 z_j \Psi(z' \Lambda z)) = \omega_{1j}^b \quad ; i \neq j \quad 6A.3$$

$$c) \quad E(z_1^2 z_j^2 \Psi(z' \Lambda z)) = \omega_{ij}^c \quad ; i \neq j \quad 6A.4$$

$$d) \quad E(z_1^2 z_j^2 z_k \Psi(z' \Lambda z)) = \omega_{ijk}^d \quad ; i \neq j, k \quad j \neq k \quad 6A.5$$

$$e) \quad E(z_1 z_j z_k z_\ell \Psi(z' \Lambda z)) = \omega_{ijkl}^e \quad ; i \neq j, k, \ell, \quad j \neq k, \ell, \quad k \neq \ell \quad 6A.6$$

Proof.

$$\begin{aligned}
 \text{a)} \quad E(z_i^4 \Psi(z' \Lambda z)) &= E \left[E(z_i^4 I(\lambda_i z_i^4 + \sum_{h \neq i} \lambda_h z_h^2)) \right] \\
 &= E \left[3\Psi(\lambda_i \chi_{(5; \theta_i)}^{2'} + \sum_{h \neq i} \lambda_h z_h^2) + 6\zeta_i^2 \Psi(\lambda_i \chi_{(7; \theta_i)}^{2'} \right. \\
 &\quad \left. + \sum_{h \neq i} \lambda_h z_h^2) + \zeta_i^4 \Psi(\lambda_i \chi_{(9; \theta_i)}^{2'} + \sum_{h \neq i} \lambda_h z_h^2) \right] = 3P_{5i} + 6\zeta_i^2 P_{7i} + \zeta_i^4 P_{9i}
 \end{aligned}$$

by Lemma 1 of Clarke *et al.* (1987a).

$$\begin{aligned}
 \text{b)} \quad E(z_i^3 z_j \Psi(z' \Lambda z)) &= E \left[E(z_i^3 E[z_j \Psi(\lambda_i z_i^2 + \lambda_j z_j^2 + \sum_{h \neq i, j} \lambda_h z_h^2)]) \right] ; i \neq j \\
 &= E \left[E(z_i^3 \zeta_j \Psi(\lambda_i z_i^2 + \lambda_j \chi_{(3; \theta_j)}^{2'} + \sum_{h \neq i, j} \lambda_h z_h^2)) \right] ; i \neq j \\
 &= E \left[\zeta_j \left(3\zeta_i \Psi(\lambda_i \chi_{(5; \theta_i)}^{2'} + \lambda_j \chi_{(3; \theta_j)}^{2'} + \sum_{h \neq i, j} \lambda_h z_h^2) \right. \right. \\
 &\quad \left. \left. + \zeta_i^3 \Psi(\lambda_i \chi_{(2; \theta_i)}^{2'} + \lambda_j \chi_{(3; \theta_j)}^{2'} + \sum_{h \neq i, j} \lambda_h z_h^2) \right) \right] \\
 &= 3\zeta_i \zeta_j P_{5i, 3j} + \zeta_i^3 \zeta_j P_{7i, 3j} ; i \neq j
 \end{aligned}$$

by Lemma 2 of Judge and Bock (1978, p.320) and equation 6A.1.

$$\begin{aligned}
 \text{c)} \quad E(z_i^2 z_j^2 \Psi(z' \Lambda z)) &= E \left[E(z_i^2 E(z_j^2 \Psi(\lambda_i z_i^2 + \lambda_j z_j^2 + \sum_{h \neq i, j} \lambda_h z_h^2))) \right] ; i \neq j \\
 &= E \left[E(z_i^2 \Psi(\lambda_i z_i^2 + \lambda_j \chi_{(3; \theta_j)}^{2'} + \sum_{h \neq i, j} \lambda_h z_h^2)) \right. \\
 &\quad \left. E(z_j^2 \zeta_i^2 \Psi(\lambda_i z_i^2 + \lambda_j \chi_{(5; \theta_j)}^{2'} + \sum_{h \neq i, j} \lambda_h z_h^2)) \right] ; i \neq j \\
 &= E \left[\Psi(\lambda_i \chi_{(3; \theta_i)}^{2'} + \lambda_j \chi_{(3; \theta_j)}^{2'} + \sum_{h \neq i, j} \lambda_h z_h^2) \right]
 \end{aligned}$$

$$\begin{aligned}
& + \zeta_1^2 \Psi(\lambda_1 \chi_{(5;\theta_1)}^{2'} + \lambda_j \chi_{(3;\theta_j)}^{2'}) + \sum_{h \neq 1,j} \lambda_h z_h^2) \\
& + \zeta_j^2 \Psi(\lambda_1 \chi_{(3;\theta_1)}^{2'} + \lambda_j \chi_{(5;\theta_j)}^{2'}) + \sum_{h \neq 1,j} \lambda_h z_h^2) \\
& + \zeta_1^2 \zeta_j^2 \Psi(\lambda_1 \chi_{(5;\theta_1)}^{2'} + \lambda_j \chi_{(5;\theta_j)}^{2'}) + \sum_{h \neq 1,j} \lambda_h z_h^2) \Big] \quad ; \quad i \neq j \\
& = P_{31,3j} + \zeta_1^2 P_{51,3j} + \zeta_j^2 P_{31,5j} + \zeta_1^2 \zeta_j^2 P_{51,5j} \quad ; \quad i \neq j
\end{aligned}$$

by Lemma 1 of Judge and Bock (1978, p.32).

Similarly we can show

$$\begin{aligned}
d) \quad E(z_1^2 z_j z_k \Psi(z' \Lambda z)) &= E_{h \neq 1,j,k} \left[E_{i=1} \left[z_i^2 E_j \left[z_j E_k \left[z_k \Psi(\lambda_1 z_1^2 + \lambda_j z_j^2 + \lambda_k z_k^2 + \sum_{h \neq 1,j,k} \lambda_h z_h^2) \right] \right] \right] \right] \\
&= \zeta_j \zeta_k P_{31,3j,3k} + \zeta_1^2 \zeta_j \zeta_k P_{51,3j,3k} \quad ; \quad i = j, k, j \neq k
\end{aligned}$$

by Lemma 1 and 2 of Judge and Bock (1978, p.320), and

$$\begin{aligned}
e) \quad E(z_1 z_j z_k z_\ell \Psi(z' \Lambda z)) &= \zeta_1 \zeta_j \zeta_k \zeta_\ell P_{31,3j,3k,3\ell} \quad ; \quad i = j, k, \ell, \\
&\quad j \neq k, \ell, k \neq \ell
\end{aligned}$$

by Lemma 2 of Judge and Bock (1978, p.320). □

$$\begin{aligned}
\text{Now } z' Vz &= \sum_{i=1}^T \sum_{j=1}^T z_j V_{ji} z_i, \text{ similarly } z' Wz = \sum_{k=1}^T \sum_{\ell=1}^T z_\ell W_{\ell k} z_k, \text{ therefore} \\
E[z' Vzz' Wz \Psi(z' \Lambda z)] &= \sum_i \sum_j \sum_k \sum_\ell \left(V_{ji} W_{\ell k} E(z_1 z_j z_k z_\ell \Psi(z' \Lambda z)) \right).
\end{aligned}$$

Hence,

$$\begin{aligned}
E(z' Vzz' Wz) &= \left(\sum_i V_{ii} W_{ii} \omega_i^a \right) + \left(\sum_i \sum_{\ell \neq 1} V_{ii} W_{\ell 1} \omega_{1\ell}^b \right) + \left(\sum_i \sum_{k \neq 1} V_{ii} W_{ik} \omega_{ik}^b \right) \\
&+ \left(\sum_i \sum_{j \neq 1} V_{ji} W_{ii} \omega_{ij}^b \right) + \left(\sum_i \sum_{j \neq 1} V_{ji} W_{jj} \omega_{ji}^b \right) + \left(\sum_i \sum_{k \neq 1} V_{ii} W_{kk} \omega_{ik}^c \right) \\
&+ \left(\sum_i \sum_{j \neq 1} V_{ji} W_{ji} \omega_{ij}^c \right) + \left(\sum_i \sum_{j \neq 1} V_{ji} W_{ij} \omega_{ij}^c \right) + \left(\sum_i \sum_{k \neq 1} \sum_{\ell \neq 1, h} V_{ii} W_{\ell k} \omega_{ik\ell}^d \right)
\end{aligned}$$

$$\begin{aligned}
& + \left(\sum_i \sum_{j \neq i} \sum_{\ell \neq i, j} V_{ji} W_{\ell i} \omega_{1j\ell}^d \right) + \left(\sum_i \sum_{j \neq i} \sum_{k \neq i} V_{ji} W_{ik} \omega_{1jk}^d \right) + \left(\sum_i \sum_{j \neq i} \sum_{k \neq i, j} V_{ji} W_{jk} \omega_{jlk}^d \right) \\
& + \left(\sum_i \sum_{j \neq i} \sum_{k \neq i, j} V_{ji} W_{kk} \omega_{kij}^d \right) + \left(\sum_i \sum_j \sum_{\ell \neq i, j} V_{ji} W_{\ell j} \omega_{jil}^d \right) \\
& + \left(\sum_i \sum_{j \neq i} \sum_{k \neq i, j} \sum_{\ell \neq i, j, k} V_{ji} W_{\ell k} \omega_{ijk\ell}^e \right).
\end{aligned}$$

Because V and W are symmetric this expression can be simplified

$$\begin{aligned}
E(z' V z z' W z) = & \sum_i \left[V_{ii} W_{ii} \omega_i^a + \sum_{j \neq i} [2V_{ii} W_{ji} \omega_{ij}^b + 2V_{ji} W_{ii} \omega_{ij}^b \right. \\
& + V_{ii} W_{jj} \omega_{ij}^c + 2V_{ji} W_{ji} \omega_{ij}^c + \sum_{k \neq i, j} [V_{ii} W_{jk} \omega_{ijk}^d + 2V_{ji} W_{ki} \omega_{ijk}^d \\
& \left. + 2V_{ij} W_{ik} \omega_{ijk}^d + V_{jk} W_{ii} \omega_{ijk}^d + \sum_{\ell = i, j, k} V_{ji} W_{\ell k} \omega_{ijk\ell}^e] \right].
\end{aligned}$$

□

Proof of Theorem 6.3.

$$\begin{aligned}
\hat{\sigma}_1^2 = & \sigma^2 \left((T+\Delta)^{-1} z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z + (T+\Gamma)^{-1} \left((T+\Delta)^{-1} (\Delta - \Gamma) z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z \right. \right. \\
& \left. \left. + z' \Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon z \right) \Psi(z' \Lambda z) \right)
\end{aligned}$$

hence

$$\begin{aligned}
E[(\hat{\sigma}_1^2)^2] = & \sigma^4 \left((T+\Delta)^{-2} (z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z)^2 \right. \\
& + 2(T+\Delta)^{-2} (T+\Gamma)^{-1} (\Delta - \Gamma) (z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z)^2 \Psi(z' \Lambda z) \\
& + 2(T+\Delta)^{-1} (T+\Gamma)^{-1} (z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z z' \Upsilon' \Omega^{1/2} X C X \Omega^{1/2} \Upsilon z) \Psi(z' \Lambda z) \\
& + (T+\Delta)^{-2} (T+\Gamma)^{-2} (\Delta - \Gamma)^2 (z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z)^2 \Psi(z' \Lambda z) \\
& + 2(T+\Delta)^{-1} (T+\Gamma)^2 (\Delta - \Gamma) (z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z z' \Upsilon' \Omega^{1/2} X C X \Omega^{1/2} \Upsilon z) \Psi(z' \Lambda z) \\
& \left. + (T+\Gamma)^{-2} (z' \Upsilon' \Omega^{1/2} X C X \Omega^{1/2} \Upsilon z)^2 \Psi(z' \Lambda z) \right)
\end{aligned}$$

$$= \sigma^4 \left((T+\Delta)(T+\Gamma)^{-2} \left[(T+\Gamma)^2 (\text{tr}(M\Omega)^2 + 2\text{tr}(M\Omega)^2) \right. \right. \\ \left. \left. + (\Delta-\Gamma)(2T+\Delta+\Gamma)\Xi_1 + 2(T+\Delta)^2\Xi_2 + (T+\Delta)^{-2}\Xi_3 \right] \right)$$

by Lemma 6.1 and the definition of Ξ_1 , Ξ_2 and Ξ_3 .

Hence

$$\rho(\hat{\sigma}_1^2, \sigma^2) = \left((T+\Delta)(T+\Gamma)^{-2} \left[(T+\Gamma)^2 (\text{tr}(M\Omega)^2 + 2\text{tr}(M\Omega)^2) (\Delta+\Gamma)(2T+\Delta+\Gamma)\Xi_1 + 2(T+\Delta)^2\Xi_2 \right. \right. \\ \left. \left. + (T+\Delta)^2\Xi_3 \right] - 2 \left[(T+\Delta)^{-1}\text{tr}(M\Omega) + (T+\Delta)^{-1}(T+\Gamma)^{-1}(\Delta-\Gamma)\text{tr}(M\Omega^{1/2}\Upsilon\Upsilon'\Omega^{1/2}) \right. \right. \\ \left. \left. + (T+\Gamma)^{-1}\text{tr}(XCX'\Omega^{1/2}\Upsilon\Upsilon'\Omega^{1/2}) \right] \right) + 1$$

by Lemma 4A.1. □

Proof of Corollary 6.2.

As $c \rightarrow 0$ the probabilities, the P_{ij} 's, all tend towards zero. Therefore P_3 , B , ω_1^a , ω_{1j}^b , ω_{1j}^c , ω_{1jk}^d and ω_{1jkl}^d tend towards zero also.

$$\begin{aligned} \text{a) } \lim_{c \rightarrow 0} \left[B(\hat{\sigma}_1^2, \sigma^2) \right] &= (T+\Delta)^{-1} \text{tr}(M\Omega) - 1 = B(s_1^2, \sigma^2). \\ \text{b) } \lim_{c \rightarrow 0} \left[\rho(\hat{\sigma}_1^2, \sigma^2) \right] &= (T+\Delta)^{-2} \left[(\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 \right] - 2(T+\Delta)^{-1} \text{tr}(M\Omega) + 1 \\ &= \rho(\hat{\sigma}_1^2, \sigma^2). \end{aligned}$$

As $c \rightarrow \infty$ the probabilities tend towards unity and hence

$$\begin{aligned} \lim_{c \rightarrow \infty} (B) &= I_T + \zeta\zeta', \\ \lim_{c \rightarrow \infty} (\omega_1^a) &= 3 + 6\zeta_1^2 + \zeta_1^4, \\ \lim_{c \rightarrow \infty} (\omega_{1j}^b) &= 3\zeta_1\zeta_j + \zeta_1^3\zeta_j, \\ \lim_{c \rightarrow \infty} (\omega_{1j}^c) &= 1 + \zeta_1^2 + \zeta_j^2 + \zeta_1 + \zeta_j^2, \end{aligned}$$

$$\lim_{c \rightarrow \infty} (\omega_{ijk}^d) = \zeta_j \zeta_k + \zeta_1^2 \zeta_j \zeta_k,$$

and

$$\lim_{c \rightarrow \infty} (\omega_{ijk\ell}^e) = \zeta_j \zeta_j \zeta_k \zeta_\ell.$$

Hence

$$\begin{aligned} \text{c) } \lim_{c \rightarrow \infty} \left(B(\hat{\sigma}_1^2, \sigma^2) \right) &= (T+\Delta)^{-1} \text{tr}(M\Omega) + (T+\Gamma)^{-1} \left[(T+\Delta)^{-1} (\Delta - \Gamma) (M\Omega^{1/2} \Upsilon(I_T + \zeta \zeta') \Upsilon \Omega^{1/2}) \right. \\ &\quad \left. + \text{tr}(XCX' \Omega^{1/2} \Upsilon(I_T + \zeta \zeta') \Upsilon \Omega^{1/2}) \right] - 1 \\ &= (T+\Delta)^{-1} \text{tr}(M\Omega) + (T+\Gamma)^{-1} \left((T+\Delta)^{-1} (\Delta - \Gamma) \text{tr}(M\Omega) + \text{tr}(XCX' \Omega) \right. \\ &\quad \left. + \zeta' \Upsilon' \Omega^{1/2} XCX' \Omega^{1/2} \Upsilon \zeta \right) - 1 \\ &= (T+\Gamma)^{-1} \left((T+\Delta)^{-1} (T+\Gamma) \text{tr}(M\Omega) + (T+\Delta)^{-1} (\Delta - \Gamma) \text{tr}(M\Omega) \text{tr}(XCX' \Omega) \right. \\ &\quad \left. + \sigma \delta' \eta' X' \Omega^{-1/2} \Upsilon \Upsilon' XCX' \Omega^{1/2} \Upsilon \Upsilon' \Omega^{-1/2} X \eta \delta \right) - 1 \\ &= (T+\Gamma)^{-1} \left(\text{tr}(XCX' \Omega + M\Omega) + \sigma^{-2} \delta' (RS^{-1}R')^{-1} \delta \right) - 1 \\ &= B(s_1^{*2}, \sigma^2). \end{aligned}$$

Now consider

$$\begin{aligned} \lim_{c \rightarrow \infty} E[z' V z z' W z \Psi(z' \Lambda z)] &= \sum_{l=1}^T \left[V_{ll} W_{ll} (3 + 6\zeta_1^2 + \zeta_1^4) \right. \\ &\quad + \sum_{j \neq l}^T \left[(V_{ll} W_{jl} + V_{jl} W_{ll}) 2(3\zeta_1 \zeta_j + \zeta_1^3 \zeta_j) + (V_{ll} W_{jj} + 2V_{jl} W_{jl}) (1 + \zeta_1^2 + \zeta_j^2 + \zeta_1^2 \zeta_j^2) \right. \\ &\quad + \sum_{k \neq j, l}^T \left[(V_{ll} W_{jk} + 2V_{jl} W_{kl} + 2V_{lj} W_{lk} + V_{jk} W_{ll}) (\zeta_j \zeta_k + \zeta_1^2 \zeta_j \zeta_k) \right. \\ &\quad \left. \left. + \sum_{\ell \neq l, j, k}^T V_{jl} W_{\ell k} (\zeta_1 \zeta_j \zeta_k \zeta_\ell) \right] \right] \right]. \end{aligned}$$

Note that, for any $T \times T$ symmetric matrices D and E and $T \times 1$ vector a ,

$$a' D a a' E a = \sum_i \sum_j \sum_k \sum_\ell D_{ji} E_{\ell k} a_i a_j a_k a_\ell$$

$$\begin{aligned}
&= \sum_i D_{ii} E_{ii} a_i^4 + \sum_{j \neq i} (D_{ii} E_{ji} + D_{ji} E_{ii}) a_i^3 a_j \\
&\quad + (D_{ii} E_{jj} + 2D_{ji} E_{ji}) a_i^2 a_j^2 \\
&\quad + \sum_{k \neq i, j} (D_{ii} E_{jk} + 2D_{ji} E_{ki} + 2D_{ji} E_{kl} + D_{jk} E_{ii}) a_i^2 a_j a_k \\
&\quad + \sum_{\ell \neq i, j, k} D_{ji} E_{\ell k} a_i a_j a_k a_\ell,
\end{aligned}$$

$$\begin{aligned}
a' DEa &= \sum_i \sum_j \sum_k D_{jk} E_{ki} a_i a_j \\
&= \sum_i D_{ii} E_{ii} a_i^2 + \sum_{j \neq i} D_{ji} E_{ji} a_i^2 + D_{ji} E_{ii} a_i a_j + D_{jj} E_{ji} a_i a_j + \sum_{k \neq i, j} D_{ji} E_{ik} a_i a_j a_k,
\end{aligned}$$

$$\begin{aligned}
a' \text{Datr}(E) &= \sum_i \sum_j \sum_k D_{ji} E_{jk} a_i a_j \\
&= \sum_i D_{ii} E_{ii} a_i^2 + \sum_{j \neq i} D_{ii} E_{jj} a_i^2 + 2D_{ji} E_{ii} a_i a_j + \sum_{k \neq i, j} D_{jk} E_{ii} a_i a_j a_k,
\end{aligned}$$

$$\text{tr}(DE) = \sum_i \sum_j D_{ij} E_{ji} = \sum_i D_{ii} E_{ii} + \sum_{j \neq i} D_{ji} E_{ji},$$

and

$$\text{tr}(D)\text{tr}(E) = \sum_i \sum_j D_{ii} E_{jj} = \sum_i D_{ii} E_{ii} + \sum_{j \neq i} D_{ii} E_{jj}.$$

Hence

$$\begin{aligned}
\lim_{c \rightarrow \infty} E[z' V z z' W z \Psi(z' \Lambda z)] &= \zeta' V \zeta \zeta' W \zeta + 4\zeta' V W \zeta + \zeta' V \zeta \text{tr}(W) \\
&\quad + \zeta' W \zeta \text{tr}(V) + 2\text{tr}(VW) + \text{tr}(V)\text{tr}(W).
\end{aligned}$$

Therefore

$$\lim_{c \rightarrow \infty} (\Xi_1) = 2\text{tr}(M\Omega)^2 + (\text{tr}(M\Omega))^2,$$

$$\lim_{c \rightarrow \infty} (\Xi_2) = \sigma^{-2} \delta' (RS^{-1}R')^{-1} \delta \text{tr}(M\Omega) + 2\text{tr}(M\Omega X C X' \Omega) + \text{tr}(M\Omega) \text{tr}(X C X' \Omega)$$

and

$$\lim_{c \rightarrow \infty} (\Xi_3) = \frac{1}{\sigma^4} (\delta' (RS^{-1}R')^{-1} \delta)^2 + \frac{4}{\sigma^2} \delta' \eta' X \Omega X \eta \delta + \frac{2}{\sigma^2} \delta' (RS^{-1}R')^{-1} \text{tr}(X C X' \Omega)$$

$$+ 2\text{tr}(XCX'\Omega)^2 + (\text{tr}(XCX'\Omega))^2,$$

as

$$\zeta' \Upsilon' \Omega^{1/2} M = \sigma^{-1} \delta' (RS^{-1}R')^{-1} RS^{-1} X' \Omega^{-1/2} \Upsilon \Upsilon' \Omega^{1/2} M = 0$$

and

$$\begin{aligned} \zeta' \Upsilon' \Omega^{1/2} XCX' \Omega^{1/2} \Upsilon \zeta &= \sigma^2 \delta' (RS^{-1}R')^{-1} RS^{-1} X' \Omega^{-1/2} \Upsilon \Upsilon' \Omega^{1/2} X S^{-1} R' \\ &\quad (RS^{-1}R')^{-1} RS^{-1} X' \Omega^{1/2} \Upsilon \Upsilon' \Omega^{-1/2} X S^{-1} R' (RS^{-1}R')^{-1} \delta \\ &= \sigma^{-2} \delta' (RS^{-1}R')^{-1} \delta. \end{aligned}$$

Thus

$$\begin{aligned} \lim_{c \rightarrow \infty} \left[\rho(\hat{\sigma}_1^2, \sigma^2) \right] &= ((T+\Delta)(T+\Gamma))^{-2} \left[(T+\Gamma)^{-2} \left((\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 \right) \right. \\ &\quad + (\Delta-\Gamma)(2T+\Delta+\Gamma) \left((\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 \right) \\ &\quad + 2(T+\Delta)^2 \left(\sigma^{-2} (RS^{-1}R')^{-1} \delta \text{tr}(M\Omega) + 2\text{tr}(M\Omega XCX' \Omega) \right. \\ &\quad \left. \left. + \text{tr}(M\Omega) \text{tr}(XCX' \Omega) \right) \right. \\ &\quad + (T+\Delta)^2 \left[\sigma^{-4} \left(\delta' (RS^{-1}R')^{-1} + 4\sigma^{-2} \delta' \eta' X' \Omega X \eta \delta \right) \right. \\ &\quad \left. \left. + 2\sigma^{-2} \delta' (RS^{-1}R')^{-1} \delta \text{tr}(XCX' \Omega) + 2\text{tr}(XCX' \Omega)^2 + (\text{tr}(XCX' \Omega))^2 \right) \right. \\ &\quad \left. - 2 \left[(T+\Delta)^{-1} \text{tr}(M\Omega) + (T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) \text{tr} \left((M\Omega^{1/2} \Upsilon (I_T + \zeta \zeta') \Upsilon' \Omega^{1/2}) \right) \right. \right. \\ &\quad \left. \left. + (T+\Gamma)^{-1} \text{tr} \left((XCX' \Omega^{1/2} \Upsilon (I_T + \zeta \zeta') \Upsilon' \Omega^{1/2}) \right) \right] + 1 \right. \\ &= (T+\Gamma)^{-2} \left[(T+\Delta)^{-2} \left((T+\Gamma)^2 + (\Delta-\Gamma)(2T+\Delta+\Gamma) \right) \left((\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 \right) \right. \\ &\quad + 2\sigma^{-2} \delta' (RS^{-1}R')^{-1} \delta \text{tr}(\phi\Omega) + 2\text{tr}(M\Omega XCX' \Omega + (XCX' \Omega)^2) \\ &\quad + 2\text{tr}(M\Omega) \text{tr}(XCX' \Omega) + (\text{tr}(XCX' \Omega))^2 + 4\sigma^{-2} \delta' \eta' X' \Omega X \eta \delta + \sigma^{-4} \left(\delta' (RS^{-1}R')^{-1} \delta \right) \left. \right] \\ &\quad - 2(T+\Gamma)^{-1} \left(\text{tr}(\phi\Omega) + \sigma^{-2} \delta' (RS^{-1}R')^{-1} \delta \right) + 1 \end{aligned}$$

$$= \rho(s_1^{*2}, \sigma^2).$$

□

Proof of Corollary 6.3.

Firstly recall that, when $\Omega = I_T$ $\Upsilon = \bar{\Upsilon}$ the joint eigenvector matrix of XCX' , $XS^{-1}X'$, M and $(XCX' - \frac{cJ}{T-K}M)$ and that

$$\lambda_i = \bar{\lambda}_i = \begin{cases} \sigma^2 & ; i = 1, \dots, J \\ 0 & ; i = J+1, \dots, K \\ -\sigma^2 \frac{cJ}{T-K} & ; i = K+1, \dots, T \end{cases}$$

and

$$\zeta_i = 0 \text{ for } i = J+1, \dots, T$$

by Lemmas 4.1 and 4.2.

Therefore, when $\Omega = I_T$,

$$\begin{aligned} B(\hat{\sigma}_1^2, \sigma^2) &= (T+\Delta)^{-1} \text{tr}(M) + (T+\Delta)^{-1} (T+\Gamma)^{-1} (\Delta-\Gamma) \text{tr}(MYBY') \\ &\quad + (T+\Gamma)^{-1} \text{tr}(XCX'YBY') - 1 \\ &= ((T+\Gamma)(T+\Delta)^{-1} ((T+\Gamma)+(\Delta-\Gamma) \text{tr}(\bar{I}_T h(0,2)) + (T+\Delta) \text{tr}(\bar{I}_J(2,0)) \\ &\quad + (T+\Delta)(\delta' \Upsilon' XCX' \Upsilon \zeta) h(4,0)) - 1 \\ &= ((T+\Gamma)(T+\Delta))^{-1} ((T+\Gamma)\nu + (\Delta-\Gamma)\nu h(0,2) + (T+\Delta)Jh(2,0) \\ &\quad + 2(T+\Delta)\lambda h(4,0)) - 1 \\ &= ((T+\Gamma)(T+\Delta))^{-1} \left((T+\Gamma) (2\lambda h(4,0) + Jh(2,0) - (T+\Gamma)(K+\Delta) \right. \\ &\quad \left. + \nu(\Delta-\Gamma)h(0,2)) \right) \end{aligned}$$

$$\text{where } \bar{I}_J = \begin{bmatrix} I_J & 0 \\ 0 & 0 \end{bmatrix}, \bar{I}_T = \begin{bmatrix} 0 & 0 \\ 0 & I_{T-K} \end{bmatrix}, \Upsilon' XCX' \Upsilon = \bar{I}_J \text{ and } \Upsilon' M \Upsilon = \bar{I}_T \text{ as } \Upsilon \text{ can be}$$

chosen to be the joint eigenvector matrix of XCX' and M when $\Omega = I_T$ and

$$\begin{aligned}
\zeta' YXCX' Y\zeta &= \sigma^{-2} \delta' (RS^{-1}R')^{-1} RS^{-1} X' Y Y' X S^{-1} R' (RS^{-1}R')^{-1} R^{-1} X' Y Y' X S^{-1} R' (RS^{-1}R')^{-1} \\
&= \sigma^{-2} \delta' (RS^{-1}R')^{-1} \delta \\
&= 2\lambda.
\end{aligned}$$

□

It is straightforward to show that, when $\Omega = I_T$,

$$\omega_1^a = 3P_{51} + 6\zeta_1^2 h(6,0) + \zeta_1^4 h(8,0),$$

$$\omega_{1j}^b = 3\zeta_1 \zeta_j h(6,0) + \zeta_1^3 \zeta_j h(8,0)$$

$$\omega_{1j}^c = P_{31,3j} + \zeta_1^2 P_{51,5j} + \zeta_j^2 P_{51,5j} + \zeta_1^2 \zeta_j^2 h(8,0)$$

$$\omega_{1jk}^d = \zeta_j \zeta_k P_{31,3j,3k} + \zeta_1^2 \zeta_j \zeta_k h(8,0)$$

and

$$\omega_{1jkl}^e = \zeta_1 \zeta_j \zeta_k \zeta_l h(8,0).$$

This follows from the fact that, when $\Omega = I_T$,

$$\begin{aligned}
P_{a1,bj,ck,d\ell} &= \Pr. \left(\lambda_1 \chi_{(a;\theta_1)}^{2'} + \lambda_j \chi_{(b;\theta_j)}^{2'} + \lambda_k \chi_{(c;\theta_1)}^{2'} + \lambda_\ell \chi_{(d;\theta_j)}^{2'} \right. \\
&\quad \left. + \sum_{h \neq 1,j,k,\ell} \lambda_h \chi_{(1;\theta_h)}^{2'} < 0 \right) \\
&= \Pr. \left(\chi_{(J+a+b+c+d-4;\lambda)}^{2'} - \frac{cJ}{T-K} \chi_{(T-K)}^2 < 0 \right) \quad ; \quad i,j,k,\ell = 1,\dots,J. \\
&= h(J+a+b+c+d-4,0) \quad ; \quad i,j,k,\ell = 1,\dots,J.
\end{aligned}$$

as $\sum_{i=1}^J \theta_i = \lambda$ and $\theta_i = 0$, for $i = K+1,\dots,T$.

It is also straightforward to show that

$$P_{a1,bj} = h(b-1,a-1) \text{ when } j = 1,\dots,J, \text{ } i = K+1,\dots,T \text{ and } \Omega = I_T,$$

as

$$P_{a1, bJ} = \Pr. \left(\lambda_j \chi_{(b; \theta_1)}^{2'} + \sum_{h \neq j}^J \lambda_h \chi_{(1; \theta_h)}^{2'} + \lambda_1 \chi_a^{2'} + \sum_{h \neq 1}^T \lambda_h \chi_1^{2'} < 0 \right)$$

$$= \Pr. \left(\chi_{(J+b-1; \lambda)}^{2'} - \frac{cJ}{T-K} \chi_{(T-K+a-1)}^2 < 0 \right)$$

for $j = 1, \dots, J$ and $i = K+1, \dots, T$.

Therefore, when $\Omega = I_T$

$$E[z' V z z' W z \Psi(z' \Lambda z)] = \sum_{i=1}^T \left[V_{ii} W_{ii} \left(3P_{51} 6\zeta_1^2 h(6,0) + \zeta_1^4 h(8,0) + \sum_{j \neq 1} \left[2(V_{ii} W_{ij} + V_{ji} W_{ii}) (3\zeta_1 \zeta_j h(6,0) + \zeta_1^3 \zeta_j h(8,0) + (V_{ii} W_{jj} + 2V_{ji} W_{ji}) (P_{31,3j} + \zeta_1^2 P_{51,3j} + \zeta_1^2 \zeta_j^2 h(8,0)) \right. \right. \right. \right.$$

$$+ \sum_{k \neq 1, j} \left[(V_{ii} W_{jk} + 2V_{ji} W_{ki} + 2V_{ij} W_{ik} + V_{jk} W_{ii}) (\zeta_j \zeta_k P_{31,3j,3k} + \zeta_1^2 \zeta_j \zeta_k h(8,0)) + \sum_{\ell \neq 1, j, k} V_{ij} W_{\ell k} \zeta_1 \zeta_j \zeta_k \zeta_\ell h(8,0) \right]$$

$$= \zeta' V \zeta \zeta' W \zeta h(8,0) + \sum_{i=1}^T \left[V_{ii} V_{ii} (3P_{51} + 6\zeta_1^2 h(6,0) + \sum_{j \neq 1} \left[2(V_{ii} W_{ij} + V_{ji} W_{ii}) (3\zeta_1 \zeta_j h(6,0) + (V_{ii} W_{jj} + 2V_{ji} W_{ji}) (P_{31,3j} + \zeta_1^2 P_{51,3j} + \zeta_1^2 \zeta_j^2 h(8,0)) \right. \right. \right.$$

$$+ \sum_{k \neq 1, j, k} \left. \left. \left. (V_{ii} W_{jk} + 2V_{ji} W_{ki} + 2V_{ij} W_{ik} + V_{jk} W_{ii}) \zeta_j \zeta_k P_{31,3j,3k} \right] \right] \right].$$

When $\Omega = I_T$, $\Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon$ is a diagonal matrix, $\begin{bmatrix} 0 & 0 \\ 0 & I_{T-K} \end{bmatrix}$, as Υ forms the eigenvector matrix of M . Similarly, $\Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon = \begin{bmatrix} I_J & 0 \\ 0 & 0 \end{bmatrix}$ is also diagonal. Let $L \equiv \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon$ and $N \equiv \Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon$, $\Xi_2 = \zeta' L \zeta \zeta' N \zeta h(8,0) + \sum_{i=K+1}^T \sum_{j=1}^J L_{ii} N_{jj} (h(2,2) + \zeta_j^2 h(4,2))$

as

$$L_{ii} = \begin{cases} 0 & ; \quad i = 1, \dots, K \\ 1 & ; \quad i = K+1, \dots, T \end{cases}$$

and

$$N_{jj} = \begin{cases} 1 & ; \quad j = 1, \dots, J \\ 0 & ; \quad j = J+1, \dots, T \end{cases},$$

with

$$L_{ij} = N_{ij} = 0 \text{ for all } i \neq j.$$

$$\text{Now, as } \zeta' L = \sigma^{-1} \delta' \eta' X' \Omega^{-1/2} \Upsilon \Upsilon' \Omega^{1/2} M \Omega^{1/2} = 0 \text{ and } \sum_{i=1}^J \zeta_i^2 = 2\lambda, \quad \Xi_1 =$$

$$\nu(\nu+2)h(0,4) \text{ and } \Xi_2 = \nu h(2,2) + 2\lambda \nu h(4,2) \text{ when } \Omega = I_T.$$

Similarly,

$$\begin{aligned} \Xi_3 &= \zeta' N \zeta \zeta' N \zeta h(8,0) + \sum_{i=1}^J \left[N_{ii}^2 (3h(4,0) + 6\zeta_i^2 h(6,0)) \right. \\ &\quad \left. + \sum_{j \neq i}^J N_{ii} N_{jj} (h(4,0) + \zeta_i^2 h(6,0) + \zeta_j^2 h(6,0)) \right] \\ &= \frac{1}{\sigma^4} (\delta' (RS^{-1}R')^{-1} \delta)^2 h(8,0) + \sum_{i=1}^J (3h(4,0) + 6\zeta_i^2 h(6,0) + (J-1) (h(4,0) \\ &\quad + \zeta_i^2 h(6,0)) + (2\lambda_1 - \zeta_i^2) h(6,0)) \\ &= 4\lambda_1^2 h(8,0) + J(J+2)h(4,0) + 4\lambda_1 (J+2)h(6,0). \end{aligned}$$

Therefore, when $\Omega = I_T$,

$$\begin{aligned} \rho(\hat{\sigma}_1^2, \sigma^2) &= ((T+\Delta)(T+\Gamma))^{-2} \left[(T+\Gamma)^2 \nu(\nu+2) + (\Delta-\Gamma)(2T+\Delta+\Gamma) \nu(\nu+2) h(0,4) \right. \\ &\quad \left. + 2(T+\Delta)^2 (\nu h(2,2) + 2\lambda \nu h(4,2)) + (T+\Delta)^2 (J(J+2)h(4,0) \right. \\ &\quad \left. + 4\lambda(J+2)h(6,0) + 4\lambda_1^2 h(8,0)) - 2 \left[(T+\Delta)(T+\Gamma)^2 \nu + (T+\Delta)(T+\Gamma)(\Delta-\Gamma) \nu h(0,2) \right. \right. \\ &\quad \left. \left. + (T+\Delta)^2 (T+\Gamma) (Jh(2,0) + 2\lambda h(4,0)) \right] \right] + 1 \\ &= ((T+\Delta)(T+\Gamma))^{-2} \left[\nu(\nu+2)(\Delta-\Gamma)(2T+\Delta+\Gamma) h(0,4) + \nu(\nu+2)(T+\Gamma)^2 \right. \\ &\quad \left. + 4\lambda(T+\Delta)^2 \nu h(4,2) + J(T+\Delta)^2 2\nu h(2,2) + 4\lambda(T+\Delta)^2 (\lambda h(8,0) + (J+2)h(6,0)) \right] \end{aligned}$$

$$\begin{aligned}
& + J(T+\Delta)^2(J+2)h(4,0) - 2(T+\Gamma)(T+\Delta)(T+\Gamma)\nu - 2(T+\Gamma)(T+\Delta)\nu(\Delta-\Gamma)h(0,2) \\
& - 2(T+\Gamma)(T+\Delta)J(T+\Delta)h(2,0) - 4\lambda(T+\Delta)^2(T+\Gamma)h(4,0) \Big] + 1. \quad \square
\end{aligned}$$

Proof of Lemma 6.3.

Recall

$$X \equiv \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \text{ and } R = \begin{bmatrix} I_J & -I_J \end{bmatrix}.$$

Therefore

$$M = \begin{bmatrix} I_{T_2} - X_1 S_1^{-1} X_1' & 0 \\ 0 & I_{T_2} - X_2 S_2^{-1} X_2' \end{bmatrix} = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}$$

where $S_1 = X_1' X_1$ and $M_1 = I_{T_1} - X_1 S_1^{-1} X_1'$

and

$$XCX' = \begin{bmatrix} X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' & -X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_2^{-1} X_2' \\ -X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' & X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_2^{-1} X_2' \end{bmatrix}.$$

Let v_1^a , $i = 1, \dots, J$, denote the $\frac{1}{2}K = J$ non-zero eigenvectors of $\Omega^{1/2}XCX'\Omega^{1/2}$ and let Υ^a denote the $T \times J$ matrix with i 'th column v_1^a . Similarly, let the J zero eigenvectors of $(\Omega^{1/2}XCX'\Omega^{1/2} - \Omega^{1/2}M\Omega^{1/2})$ be denoted v_1^b ; $i = J+1, \dots, K$, with corresponding $T \times J$ matrix Υ^b . Let the zero eigenvectors of $X_1 S_1^{-1} X_1'$ be denoted v_1^c ; $i = K+1, \dots, \frac{1}{2}(T+K)$ with corresponding matrix $T_1 \times (T_1 - K_1) \Upsilon^c$ and let the zero eigenvectors of $X_2 S_2^{-1} X_2'$ be denoted v_1^d ; $i = \frac{1}{2}(T+K)+1, \dots, T$, with corresponding $T_2 \times (T_2 - K_2)$ eigenvector matrix v^d .

Consider the $T \times T$ matrix $\tilde{\Upsilon}$ with i 'th column denoted \tilde{v}_1 where

$$\tilde{v}_1 = \begin{cases} v_1^a & ; i = 1, \dots, J \\ v_1^b & ; i = J+1, \dots, K \\ \begin{bmatrix} v_1^c \\ 0 \end{bmatrix} & ; i = K+1, \dots, \frac{1}{2}(T+K) \\ \begin{bmatrix} 0 \\ v_1^d \end{bmatrix} & ; i = \frac{1}{2}(T+K)+1, \dots, T \end{cases}$$

To show that \tilde{v} is the joint eigenvector matrix of $\Omega^{1/2}M\Omega^{1/2}$, $\Omega^{1/2}XCX'\Omega^{1/2}$ and $\Phi = \Omega^{1/2}(XCX' - \frac{cJ}{T-K}M)$ it is sufficient to show that each \tilde{v}_1 is an eigenvector of both $\Omega^{1/2}M\Omega^{1/2}$ and $\Omega^{1/2}XCX'\Omega^{1/2}$ and that $\tilde{v}_i' \tilde{v}_j = 0$ for all $i \neq j$. We already have $\tilde{v}_i' \tilde{v}_i = 1$ by definition.

a) $\Omega^{1/2}XCX'\Omega^{1/2}\tilde{v}_1 = \Omega^{1/2}XCX'\Omega^{1/2}v_1^a = \lambda_1^c v_1^a$ with $\lambda_1^c \neq 0$; $i = 1, \dots, J$, by definition.

$$\begin{aligned} & \text{Note that } \Omega^{1/2}M\Omega^{1/2}\Omega^{1/2}XCX'\Omega^{1/2} \\ &= \begin{bmatrix} M_1 & 0 \\ 0 & \psi M_2 \end{bmatrix} \begin{bmatrix} X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' & -\psi^{1/2} X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_2^{-1} X_2' \\ -\psi^{1/2} X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' & \psi X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_2^{-1} X_2' \end{bmatrix} \\ &= \begin{bmatrix} M_1 X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' & -\psi^{1/2} M_1 X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_2^{-1} X_2' \\ -\psi^{3/2} M_2 X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' & \psi^2 X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_2^{-1} X_2' \end{bmatrix} \\ &= 0 \end{aligned}$$

Now $\Omega^{1/2}XCX'\Omega^{1/2}\tilde{v}_1 = \Omega^{1/2}XCX'\Omega^{1/2}v_1^a = \lambda_1^c \tilde{v}_1$; $i = 1, \dots, J$, by definition, hence $\Omega^{1/2}M\Omega^{1/2}\Omega^{1/2}XCX'\Omega^{1/2}\tilde{v}_1 = \lambda_1^c \Omega^{1/2}M\Omega^{1/2}\tilde{v}_1 = 0$. Therefore each \tilde{v}_i ; $i = 1, \dots, J$ is a non-zero eigenvector of $\Omega^{1/2}XCX'\Omega^{1/2}$ and a zero eigenvector of $\Omega^{1/2}M\Omega^{1/2}$.

b) $(\Omega^{1/2}XCX'\Omega^{1/2}\tilde{v}_1 - \Omega^{1/2}M\Omega^{1/2})\tilde{v}_1 = 0$; $i = J+1, \dots, K$ by definition.

Hence

$$(\Omega^{1/2}M\Omega^{-1/2}\Omega^{1/2}XCX'\Omega^{1/2} - \Omega^{1/2}M\Omega^{-1/2}\Omega^{1/2}M\Omega^{1/2})\tilde{v}_1 = 0; i = J+1, \dots, K$$

or

$$\Omega^{1/2} M \Omega^{1/2} \tilde{v}_1 = 0; \quad i = J+1, \dots, K.$$

Therefore $\Omega^{1/2} X C X' \Omega^{1/2} \tilde{v}_1 = 0$, $i = J+1, \dots, K$, also by definition. Hence, \tilde{v}_1 , $i = J+1, \dots, K$, is a zero eigenvector of both $\Omega^{1/2} X C X' \Omega$ and $\Omega^{1/2} M \Omega^{1/2}$.

$$\begin{aligned} \text{c) } \Omega^{1/2} M \Omega^{1/2} \tilde{v}_1 &= \begin{bmatrix} I_{T_1} - X_1 S_1^{-1} X_1' & 0 \\ 0 & \psi I_{T_2} - \psi X_2 S_2^{-1} X_2' \end{bmatrix} \begin{bmatrix} v_1^c \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} v_1^c - X_1 S_1^{-1} X_1' v_1^c \\ 0 \end{bmatrix} = \tilde{v}_1; \quad i = K+1, \dots, \frac{1}{2}(T+K) \end{aligned}$$

by definition.

Note that $X_1 S_1^{-1} X_1' v_1^c = 0$ and hence

$$X_1' X_1 S_1^{-1} X_1' v_1^c = X_1 v_1^c = 0, \quad i = K+1, \dots, \frac{1}{2}(T+K).$$

Hence

$$\begin{aligned} \Omega^{1/2} X C X' \Omega^{1/2} \tilde{v}_1 &= \begin{bmatrix} X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' & -\psi^{1/2} X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_2^{-1} X_2' \\ -\psi^{1/2} X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' & \psi X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_2^{-1} X_2' \end{bmatrix} \begin{bmatrix} v_1^c \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} X_1 S_1^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' v_1^c \\ -\psi^{1/2} X_2 S_2^{-1} (S_1^{-1} + S_2^{-1})^{-1} S_1^{-1} X_1' v_1^c \end{bmatrix} = 0; \quad i = K+1, \dots, \frac{1}{2}(T+K). \end{aligned}$$

Therefore each \tilde{v}_1 , $i = K+1, \dots, \frac{1}{2}(T+K)$, is a unit eigenvector of $\Omega^{1/2} M \Omega^{1/2}$ and a zero eigenvector of $\Omega^{1/2} X C X' \Omega^{1/2}$.

$$\text{d) } \Omega^{1/2} M \Omega^{1/2} \tilde{v}_1 = \begin{bmatrix} I_{T_1} - X_1 S_1^{-1} X_1' & 0 \\ 0 & \psi I_{T_2} - \psi X_2 S_2^{-1} X_2' \end{bmatrix} \begin{bmatrix} 0 \\ v_1^d \end{bmatrix} = \psi \begin{bmatrix} 0 \\ v_1^d \end{bmatrix}; \quad i = \frac{1}{2}(K+T)+1, \dots, T$$

by definition. Also

$$X_2 S_2^{-1} X_2' v_1^d = X_2' X_2 S_2^{-1} X_2' v_1^d = X_2' v_1^d = 0$$

hence

$$\Omega^{1/2}XCX'\Omega^{1/2}\tilde{v}_1 = \begin{bmatrix} -\psi^{1/2}X_1S_1^{-1}(S_1^{-1}+S_2^{-1})^{-1}S_2^{-1}X_2'v_1^d \\ \psi^{1/2}X_2S_2^{-1}(S_1^{-1}+S_2^{-1})^{-1}S_2^{-1}X_2'v_1^d \end{bmatrix} = 0; \quad i = \frac{1}{2}(T+K)+1, \dots, T.$$

Therefore each \tilde{v}_i , $i = \frac{1}{2}(K+K)+1, \dots, T$, is a ψ eigenvector of $\Omega^{1/2}M\Omega^{1/2}$ and a zero eigenvector of $\Omega^{1/2}XCX'\Omega^{1/2}$.

The independence of the eigenvectors follows from noting that

$$\tilde{v}_i' \tilde{v}_j = 0 \text{ for } i, j = 1, \dots, J, \quad i \neq j \text{ by definition,}$$

$$\tilde{v}_i' \tilde{v}_j = \frac{1}{\lambda_i^c} \tilde{v}_i' \Omega^{1/2}XCX'\Omega^{1/2} \tilde{v}_j = 0 \text{ for } i = 1, \dots, J, \quad j = J+1, \dots, T,$$

$$\tilde{v}_i' \tilde{v}_j = 0, \quad i, j = J+1, \dots, K, \quad i \neq j \text{ by definition,}$$

$$\tilde{v}_i' \tilde{v}_j = \tilde{v}_i' \Omega^{1/2}M\Omega^{1/2} \tilde{v}_j \frac{1}{\lambda_j^M} = 0 \text{ for } i = J+1, \dots, K, \quad j = K+1, \dots, T,$$

$$\tilde{v}_i' \tilde{v}_j = 0, \quad i, j = K+1, \dots, \frac{1}{2}(T-K), \quad i \neq j \text{ by definition.},$$

$$\tilde{v}_i' \tilde{v}_j = \begin{bmatrix} v_i^c : 0 \end{bmatrix} \begin{bmatrix} 0 \\ v_j^d \end{bmatrix} = 0 \text{ for } i = K+1, \dots, \frac{1}{2}(T+K), \quad j = \frac{1}{2}(T+K)+1, \dots, T,$$

and

$$\tilde{v}_i' \tilde{v}_j = 0, \quad i, j = \frac{1}{2}(T+K)+1, \dots, T, \quad i \neq j \text{ by definition.} \quad \square$$

CHAPTER SEVEN

PRE-TEST ESTIMATION OF THE SCALE PARAMETER IN A LINEAR MODEL WITH A MIS-SPECIFIED ERROR COVARIANCE MATRIX AND EXCLUDED RELEVANT REGRESSORS

7.1 Introduction

In the previous chapter we derived the bias and risk formulae for the unrestricted, restricted and pre-test estimators of the scale parameter in a model which is correctly specified except for the mis-specification of the error covariance matrix. In particular, the regressor matrix was well specified. In a practical situation this may not, perhaps, be very likely to occur. Due to incorrect theory, the unavailability of data, incorrect functional form, structural change or some other reason, particular relevant regressors may be effectively omitted from a linear model.

In Chapter Five we considered the consequences of such an omission in conjunction with a mis-specified error covariance matrix, on the estimators of the coefficient vector and the conditional expectation of the dependent variable. In this chapter, we consider the effects of such a mis-specification on the estimators of the scale parameter. Suppose the true data-generating process is given by

$$y = X\beta + Z\beta_z + \varepsilon ; \quad \varepsilon \sim N(0, \sigma^2 \Omega)$$

where y , X , β , Z , β_z and ε are all defined in the usual way. The restrictions to be tested are represented by the hypotheses $H_0: R\beta = r$ vs $H_A: R\beta \neq r$, with $\delta \equiv R\beta - r$. We will suppose that the regressors represented by the Z matrix are unobservable or are omitted from the fitted model for some reason or other, and that the researcher assumes that the error term is well behaved. The fitted model is, therefore, the classical

linear model described by

$$y = X\beta + \varepsilon ;$$

ε is incorrectly assumed to be $N(0, \sigma^2 I_T)$ distributed, when in fact, $\varepsilon \sim N(\xi, \sigma^2 \Omega)$ with $\xi = Z\beta_z$. The hypotheses are tested using the Wald statistic described above which, in this case, has a c.d.f. as given in Chapter Five.

The unrestricted, restricted and pre-test estimators of the scale parameter are defined as

$$s_1^2 = (y - Xb)'(y - Xb)/(T + \Delta),$$

$$s_1^{*2} = (y - Xb^*)'(y - Xb^*)/(T + \Gamma),$$

and

$$\hat{\sigma}_1^2 = \begin{cases} s_1^2 & ; \quad u \geq c \\ s_i^{*2} & ; \quad u < c \end{cases}$$

respectively, where $b = (X'X)^{-1}X'y$, $b^* = b - S^{-1}(RS^{-1}R')^{-1}(Rb - r)$, $S = X'X$, u is the test statistic, c is the critical value of the pre-test and $i = ML, LS$ and MS as $(\Delta, \Gamma) = (0, 0), (-K, J - K)$ and $(2 - K, J + 2 - K)$ respectively.

7.2 Properties of the Component Estimators

Theorem 7.1

The relative biases and risks of the unrestricted and restricted estimators are given by

$$a) \quad B(s_1^2, \sigma^2) = (T + \Delta)^{-1} \left(\text{tr}(M\Omega) + \frac{1}{\sigma^2} \xi' M \xi \right) - 1,$$

$$b) \quad \rho(s_1^2, \sigma^2) = (T + \Delta)^{-2} \left[\left(\text{tr}(M\Omega) \right)^2 + 2 \text{tr}(M\Omega)^2 + \frac{4}{\sigma^2} \xi' M \Omega M \xi + \frac{2}{\sigma^2} \xi' M \xi \text{tr}(M\Omega) + \frac{1}{\sigma^4} (\xi' M \xi)^2 \right] - 2(T + \Delta)^{-1} \left(\text{tr}(M\Omega) + \frac{1}{\sigma^2} \xi' M \xi \right) + 1,$$

$$c) \quad B(s_1^{*2}, \sigma^2) = (T + \Gamma)^{-1} \left(\text{tr}(\phi\Omega) + \frac{1}{\sigma^2} \bar{\delta}' \phi \bar{\delta} \right) - 1,$$

$$\begin{aligned}
d) \quad \rho(s_1^{*2}, \sigma^2) &= (T+\Gamma)^{-2} \left[(\text{tr}(\phi\Omega))^2 + 2\text{tr}(\phi\Omega)^2 + \frac{4}{\sigma^2} \bar{\delta}' \phi \Omega \phi \bar{\delta} + \frac{2}{\sigma^2} \bar{\delta}' \phi \bar{\delta} \text{tr}(\phi\Omega) \right. \\
&\quad \left. + \frac{1}{\sigma^4} (\bar{\delta}' \phi \bar{\delta})^2 \right] - 2(T+\Gamma)^{-1} \left(\text{tr}(\phi\Omega) + \frac{1}{\sigma^2} \bar{\delta}' \phi \bar{\delta} \right) + 1
\end{aligned}$$

respectively, where $M = I_T - XS^{-1}X'$, $\phi = (XCX' + M)$ and $\bar{\delta} = (X\eta\delta + \xi)$.

Proof.

See Appendix 7A.

It is apparent from the above formulae, that the restricted estimator is generally more biased than the unrestricted estimator with both its bias and risk increasing without bound as the hypothesis error, δ , increases. It is also apparent that the bias and risk of both estimators are unbounded as the regressor mis-specification error, ξ , increases. It is trivial to show that these formulae collapse to those given in the previous chapter for a model with no excluded regressors, by considering the case of $\xi = 0$.

Similarly we have

Corollary 7.1.

In the case of $\Omega = I_T$, the relative bias and risk of the unrestricted and restricted estimators are given by¹

- a) $B(s_1^2, \sigma^2) = (T+\Delta)^{-1} (2\lambda_2 - (K+\Delta))$
- b) $\rho(s_1^2, \sigma^2) = (T+\Delta)^{-2} (2(\nu+4\lambda_2) + (2\lambda_2 - K - \Delta)^2)$
- c) $B(s_1^{*2}, \sigma^2) = (T+\Gamma)^{-1} (J + 2(\lambda_1 + \lambda_2) - (K+\Gamma))$

¹ These risk formulae for the ML estimators are equivalent to those given by Giles and Clarke (1989) in the case of a model with excluded relevant regressors and a scalar error covariance matrix. Note that Giles and Clarke follow Mittelhammer (1984) in defining $\delta = r - R\beta$. We do not follow this convention for the sake of consistency with previous chapters of this thesis.

$$d) \quad \rho(s_1^{*2}, \sigma^2) = (T+\Gamma)^{-2} \left\{ 2(J+\nu+4(\lambda_1+\lambda_2)) + (J-K-\Gamma+2(\lambda_1+\lambda_2))^2 \right\}$$

respectively, where $\nu = T-K$, $\lambda_1 = \frac{1}{2\sigma^2}(\delta' \eta' X' + \xi')XCX'(X\eta\delta + \xi)$ and $\lambda_2 = \frac{1}{2\sigma^2}\xi' M\xi$.

Proof.

See Appendix 7A.

7.3 The Bias and Risk of the Pre-Test Estimators

Now consider the pre-test estimator

$$\hat{\sigma}_1^2 = \begin{cases} s_1^2 & \text{if } u < c \\ s_1^{*2} & \text{if } u \geq c \end{cases}$$

Recall from Chapter Five that $u < c \Leftrightarrow z' \Lambda z < 0$ where $z = \sigma^{-1} \Gamma' \Omega^{-1/2}(\epsilon + \xi + X\eta\delta)$, with $\Lambda = \text{diag}\{\lambda_i\}$ and λ_i and Γ the i 'th eigenvalue and eigenvector matrix of $\sigma^2 \Omega^{1/2}(XCX' - \frac{cJ}{T-K}M)\Omega^{1/2}$ respectively. Hence

$$\begin{aligned} \hat{\sigma}_1^2 &= (T+\Delta)^{-1}(\epsilon' M\bar{\epsilon}) + \left[(\Delta-\Gamma)(T+\Delta)^{-1}(T+\Gamma)^{-1}\bar{\epsilon}' M\bar{\epsilon} + (T+\Gamma)^{-1}\bar{\epsilon}' XCX\bar{\epsilon} \right] \Psi(z' \Lambda z) \\ &= \frac{\sigma^2}{(T+\Gamma)} z' \Gamma' \Omega^{1/2} M \Omega^{1/2} \Gamma z + \sigma^2 \left[(\Delta-\Gamma)(T+\Delta)^{-1}(T+\Gamma)^{-1} z' \Gamma \Omega^{1/2} M \Omega^{1/2} \Gamma z \right. \\ &\quad \left. + (T+\Gamma)^{-1} z' \Gamma' \Omega^{1/2} XCX' \Omega^{1/2} \Gamma z \right] \Psi(z' \Lambda z) \end{aligned}$$

Let B and $P_{m_l, n_j, o_k, p_\ell}$ be defined as in Chapter Six. It is obvious that, given the appropriate definition of $\zeta \equiv \sigma^{-1} \Gamma' \Omega^{-1/2}(\xi + X\eta\delta)$, Lemma 6.2 holds.

Theorem 7.2.

The bias and risk of the pre-test estimator are

$$\begin{aligned} a) \quad B(\hat{\sigma}_1^2, \sigma^2) &= (T+\Delta)^{-1} \left(\text{tr}(M\Omega) + \frac{1}{\sigma^2} \xi' M\xi \right) + (\Delta-\Gamma)(T+\Delta)^{-1}(T+\Gamma)^{-1} \text{tr}(M\Omega^{1/2} \Gamma B \Gamma' \Omega^{1/2}) \\ &\quad + (T+\Gamma)^{-1} \text{tr}(XCX' \Omega^{1/2} \Gamma B \Gamma' \Omega^{1/2}) - 1 \end{aligned}$$

and

$$\begin{aligned}
b) \quad \rho(\hat{\sigma}_1^2, \sigma^2) = & \left((T+\Delta)(T+\Gamma) \right)^{-2} \left[(T+\Gamma)^2 \left(\left(\text{tr}(M\Omega) \right)^2 + 2\text{tr}(M\Omega)^2 + \frac{4}{\sigma^2} \xi' M \Omega M \xi \right. \right. \\
& + \frac{2}{\sigma^2} \xi' M \xi \text{tr}(M\Omega) + \frac{1}{\sigma^4} (\xi' M \xi)^2 \Big) \\
& + (\Delta-\Gamma)(2T+\Delta+\Gamma)\Xi_1 + 2(T+\Delta)^2\Xi_2 \\
& + (T+\Delta)^2\Xi_3 \Big] - 2 \left[(T+\Delta)^{-1} \left(\text{tr}(M\Omega) + \frac{1}{\sigma^2} \xi' M \xi \right) \right. \\
& + (\Delta-\Gamma)(T+\Delta)^{-1}(T+\Gamma)^{-1} \text{tr}(M\Omega^{1/2} \Upsilon B \Upsilon' \Omega^{1/2}) \\
& \left. \left. + (T+\Gamma)^{-1} \text{tr}(X C X' \Omega^{1/2} \Upsilon B \Upsilon' \Omega^{1/2}) \right] \right] + 1,
\end{aligned}$$

where $\Xi_1 = E\left((z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z)^2 \Psi(z' \Lambda z)\right)$

$$\Xi_2 = E\left((z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z z' \Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon z)^2 \Psi(z' \Lambda z)\right)$$

$$\Xi_3 = E\left((z' \Upsilon' \Omega^{1/2} X C X' \Omega^{1/2} \Upsilon z)^2 \Psi(z' \Lambda z)\right).$$

Ξ_i , $i = 1, 2, 3$ may be evaluated using Lemma 6.2.

Proof.

See Appendix 7A.

Corollary 7.2.

As the critical value of the pre-test, c , tends towards 0 (∞), the bias and risk of the pre-test estimator tend towards the bias and risk of the unrestricted (restricted) estimator.

Proof.

See Appendix 7A.

Similarly, it is straightforward to show the following:

Corollary 7.3.

When $\Omega = I_T$ the bias and risk of the pre-test estimator are given by²

²

The formula for the pre-test risk of the ML estimator is equivalent to that given by Giles and Clarke (1989) for the case of a model with a scalar error covariance matrix and excluded relevant regressors.

$$\begin{aligned} \text{a) } B(\hat{\sigma}_1^2, \sigma^2) &= ((T+\Delta)(T+\Gamma))^{-1} \left[(T+\Gamma)(\nu+2\lambda_2) + (\Delta-\Gamma)(\nu h'(0,2) + 2\lambda_2 h'(0,4)) \right. \\ &\quad \left. + (T+\Delta)(Jh'(2,0) + 2\lambda_1 h'(4,0)) \right] - 1 \end{aligned}$$

and

$$\begin{aligned} \rho(\hat{\sigma}_1^2, \sigma^2) &= (T+\Delta)^{-2} (2(\nu+4\lambda_2) + (2\lambda_2 - K - \Delta)^2) \\ &\quad + ((T+\Delta)(T+\Gamma))^{-2} \left[(T+\Delta)^2 \left(2J\nu h'(2,2) - 2(T+\Gamma)Jh'(2,0) + 4\nu\lambda_1 h'(4,2) \right. \right. \\ &\quad + (J(J+2) - 4(T+\Gamma)\lambda_1)h'(4,0) + 4\lambda_1(J+2)h'(6,0) \\ &\quad + 4\lambda_1^2 h'(8,0) + 4\lambda_2(Jh'(2,4) + 2\lambda_1 h'(4,4)) \left. \right) \\ &\quad + (\Delta-\Gamma)(2T+\Delta+\Gamma) \left(4\lambda_2(\lambda_2 h'(0,8) + (\nu+2)h'(0,6)) \right. \\ &\quad \left. + \nu(\nu+2)h'(0,4) \right) - 2(T+\Gamma)(T+\Delta)(\Delta-\Gamma)(\nu h'(0,2) + 2\lambda_2 h'(0,4)) \left. \right] \end{aligned}$$

where

$$h'(ij) = \text{Pr.} \left[\frac{\chi_{(J+1; \lambda_1)}^{2,}}{\chi_{(\nu+J; \lambda_2)}^{2,}} < \frac{cJ}{T-K} \right].$$

Proof.

See Appendix 7A.

It is apparent that, in general, the bias and risk of the PTE and those of the unrestricted and restricted estimators, depend on the nature of both the included and excluded data, the form of restrictions, the hypothesis error and sample size.

Other than Corollaries 7.2 and 7.3, it is difficult to determine the effect of this double mis-specification on the estimators of the scale parameter without numerical evaluation. However, given the nature of the expressions, this is computationally burdensome as between T^3 and T^4 P_{ij} values must be calculated for each value of θ and each model we consider. As in Chapter Six, however, these formulae may be simplified if we consider

a special case, and it is this special case which we consider in the next section.

7.4 Numerical Evaluation

7.4.1 The Model

Consider the model

$$y \equiv \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + Z\beta_z + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \equiv X\beta + \xi + \epsilon$$

where y , y_i , X , X_i , β , β_i , ϵ and ϵ_i , $i = 1, 2$, are defined as above in Section 6.3 of Chapter Six. Let $T_1 = T_2 = \frac{1}{2}T$ and assume that $\epsilon \sim N\left(0, \sigma^2 \begin{bmatrix} I_{T_1} & 0 \\ 0 & \psi I_{T_2} \end{bmatrix}\right)$. As above, the parameter ψ measures the degree of

heteroscedasticity in the errors. The null and alternative hypotheses are $H_0: R\beta = r \Leftrightarrow \beta_1 = \beta_2$ vs $H_A: R\beta = r \Leftrightarrow \beta_1 \neq \beta_2$ with $R = [I_J; -I_J]$, $J = \frac{1}{2}K$ and $r = 0$. The null hypothesis is tested using the usual Wald statistic, defined above.

In this case the formulae describing the bias and risk of the PTE can be considerably simplified.

Corollary 7.4.

Under the assumptions of the model, in this case the bias and risk of the PTE are given by:

$$\begin{aligned} \text{a) } B(\hat{\sigma}_1^2, \sigma^2) &= ((T+\Delta)(T+\Gamma))^{-1} \left[(T+\Gamma) \left(\text{tr}(M\Omega) + \frac{1}{\sigma^2} \xi' M \xi \right) + (\Delta-\Gamma) \sum_{i=K+1}^T \lambda_i^M (P_{3i}^M + \zeta_i^2 P_{5i}^M) \right. \\ &\quad \left. + (T+\Delta) \sum_{i=1}^J \lambda_i^C (P_{3i}^C + \zeta_i^2 P_{5i}^C) \right] - 1 \end{aligned}$$

and

$$\text{b) } \rho(\hat{\sigma}_1^2, \sigma^2) = ((T+\Delta)(T+\Gamma))^{-2} \left[(T+\Gamma)^2 \left((\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 + \frac{4}{\sigma^2} \xi' M \Omega M \xi + \frac{2}{\sigma^2} \xi' M \xi \text{tr}(M\Omega) \right) \right]$$

$$\begin{aligned}
& + \frac{1}{\sigma^4} (\xi' M \xi)^2 \Big) + (\Delta - \Gamma)(2T + \Gamma + \Delta) \Bigg(\sum_{i=K+1}^T (\lambda_i^M)^2 (3P_{5i} + 6\zeta_1^2 P_{7i} + \zeta_1^4 P_{9i}) \\
& + \sum_{\substack{j=1 \\ j \neq i}}^J \lambda_i^M \lambda_j^M (P_{3i3j} + \zeta_1^2 P_{5i3j} + \zeta_j^2 P_{3i5j} + \zeta_1^2 \zeta_j^2 P_{5i5j}) \Bigg) \\
& + (T + \Delta)^2 \Bigg(2 \sum_{i=1}^J \sum_{j=K+1}^T \lambda_i^C \lambda_j^M (P_{3i3j} + \zeta_1^2 P_{5i3j} + \zeta_j^2 P_{3i5j} + \zeta_1^2 \zeta_j^2 P_{5i5j}) \\
& + \sum_{i=1}^J (\lambda_i^C)^2 (3P_{5i} + 6\zeta_1^2 P_{7i} + \zeta_1^4 P_{9i}) + \sum_{\substack{j=1 \\ j \neq i}}^J \lambda_i^C \lambda_j^C (P_{3i3j} + \zeta_1^2 P_{5i3j} + \zeta_j^2 P_{3i5j} + \zeta_1^2 \zeta_j^2 P_{5i5j}) \Bigg) \\
& - 2 \Bigg((T + \Gamma)^2 (T + \Delta) \Big(\text{tr}(M\Omega) + \frac{1}{\sigma^2} \xi' M \xi \Big) + (\Delta - \Gamma)(T + \Delta)(T + \Gamma) \sum_{i=K+1}^T \lambda_i^M (P_{3i} + \zeta_1^2 P_{5i}) \\
& + (T + \Gamma)(T + \Delta)^2 \sum_{i=1}^J \lambda_i^C (P_{3i} + \zeta_1^2 P_{5i}) \Bigg) \Bigg] + 1.
\end{aligned}$$

Proof.

The proof is essentially the same as the proof of Corollary 6.4.

These formulae have again been evaluated using a FORTRAN program, written by the author and incorporating Davies' (1980) algorithm, executed on a Vax 6340 computer. The data used are described in Appendix 4B.

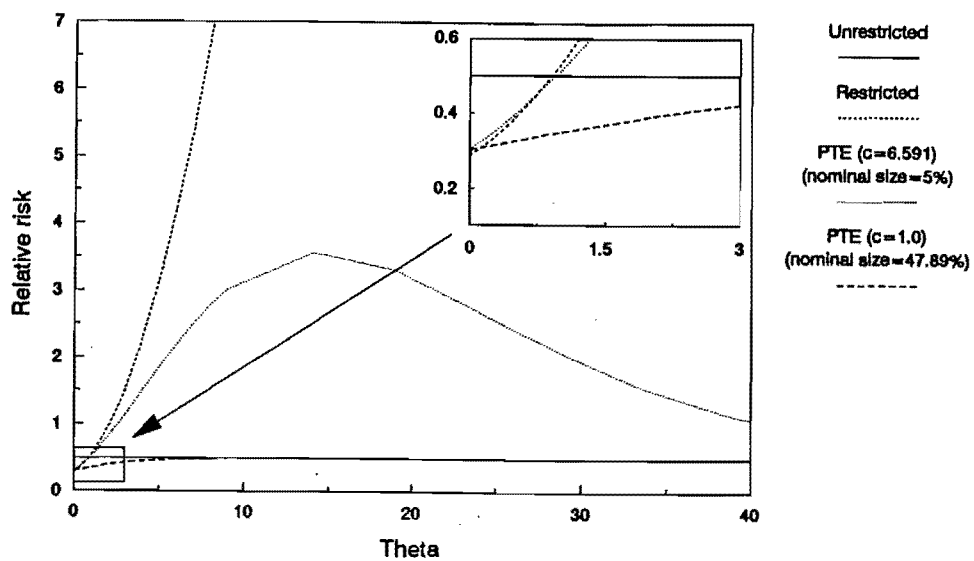
7.4.2 Numerical Results

The effects of the double mis-specification on the pre-test size and power function are described in Chapter Five. Recall that, for a given value of θ , value of $\theta_1 = \sum_{(1:\lambda_1 > 0)} \frac{1}{2\sigma^2} (v_1' \Omega^{-1/2} (X\eta\delta + \xi))^2$, any increase in the degree of regressor mis-specification, as measured by λ_2 , is associated with a decrease in the power of the pre-test. Recall also that, for a given data set, the minimum value of θ , which does not, in general, correspond to $\delta = 0$, increases as λ_2 increases. The minimum value of θ , $\bar{\theta}$, varies also with the degree of heteroscedasticity in the errors, increasing as ψ decreases and decreasing with the correlation between the included and excluded

regressors. In general the more highly correlated the included and excluded regressors are, the smaller is the value of $\bar{\theta}$ for a given value of λ_2 . Intuitively this makes sense as, if the included and excluded regressors are highly correlated, the included regressors are acting as proxies for the excluded regressors. If the included and excluded regressors were perfectly correlated we would expect the exclusions to have no effect whatsoever on our results.

Typical unrestricted, restricted and pre-test estimator risk functions for a well-specified model are shown in Figures 7.1, 7.2 and 7.3. The effects of excluding relevant regressors from the model may be seen in Figures 7.4, 7.5 and 7.6.

FIGURE 7.1: Scale Parameter Estimator Risk Fns.
 LS Estimators, Well Specified Model
 Regressors; $N(0,1)$ r.v. and linear trend

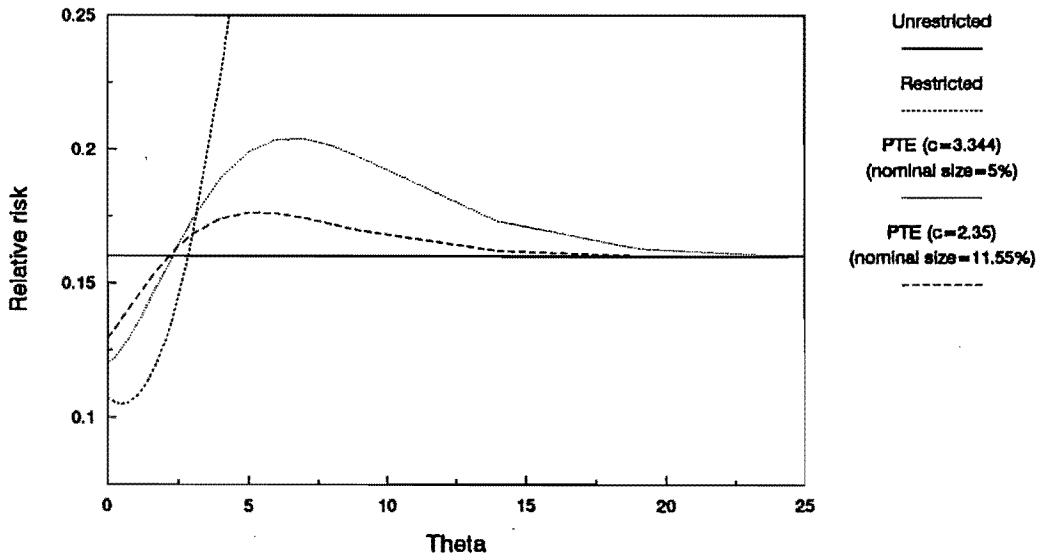


Application of Chow test

Sample size = 10

FIGURE 7.2: Scale Parameter Estimator Risk Fns.

ML Estimators, Well Specified Model
Regressors; log normal r.v. & exponential trend

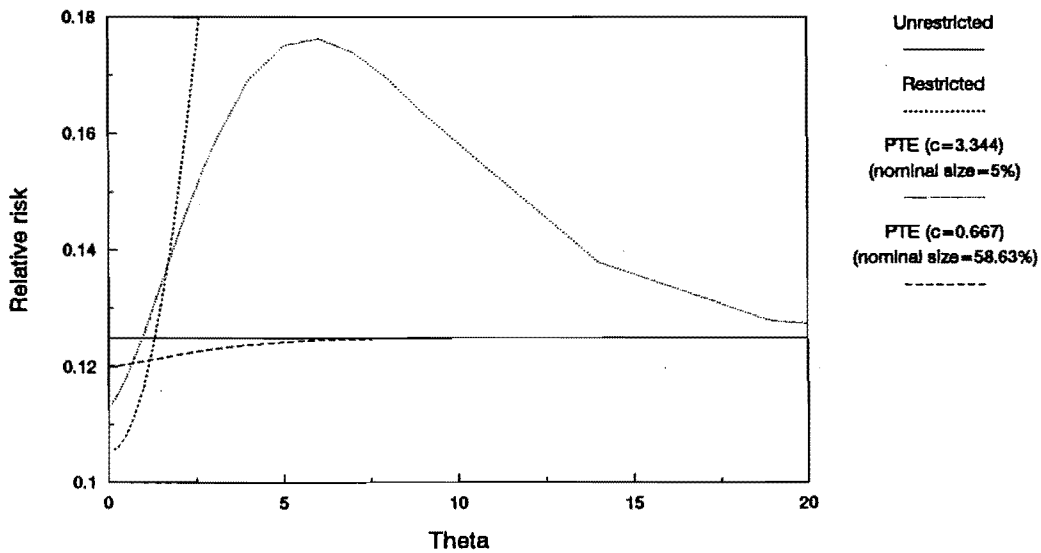


Application of Chow test

Sample size = 20

FIGURE 7.3: Scale Parameter Estimator Fns.

MS Estimators, Well Specified Model
Regressors; Aus\$-US\$ x-rate & exponential trend

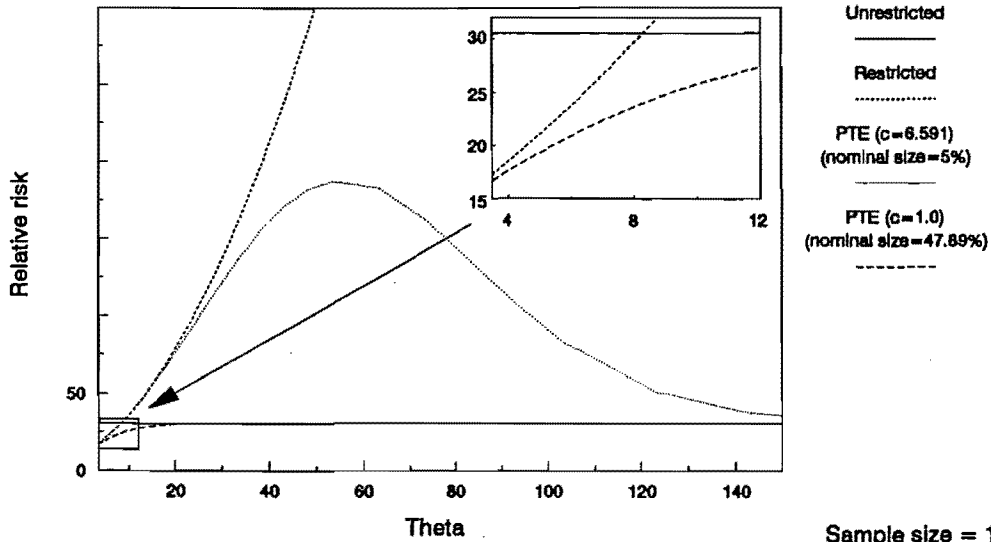


Application of Chow test

Sample size = 20

FIGURE 7.4: Scale Parameter Estimator Risk Fns.

LS Estimators, Excluded Regressors, $\Psi = 1.0$
Regressors; $N(0,1)$ r.v. and linear trend

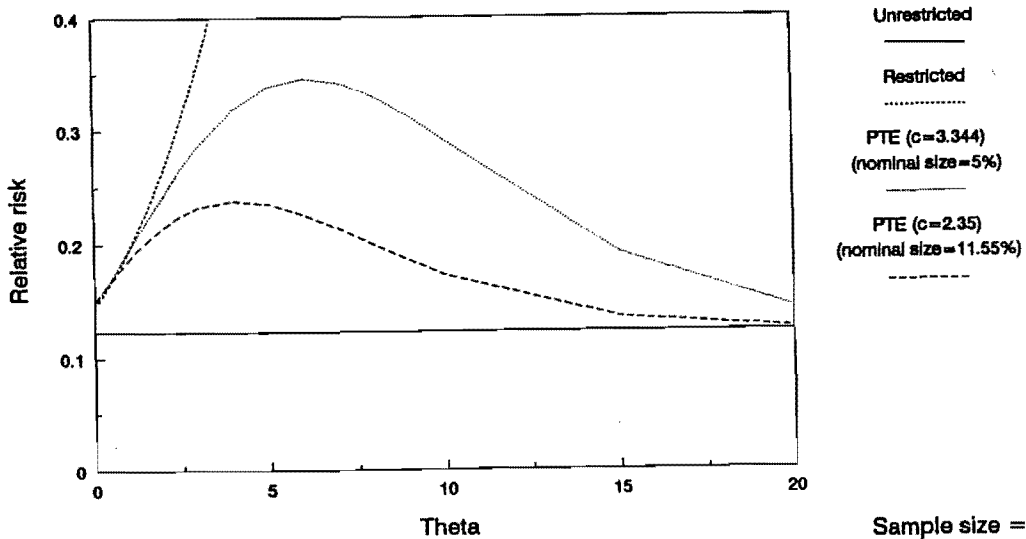


Application of Chow test
Excluded regressors are seasonal dummies

Sample size = 10
 $\lambda_2 = 10.0$
 $\theta_{\text{bar}} = 3.441$

FIGURE 7.5: Scale Parameter Estimator Risk Fns.

ML Estimators, Excluded Regressor, $\Psi = 1.0$
Regressors; log normal r.v. & exponential trend

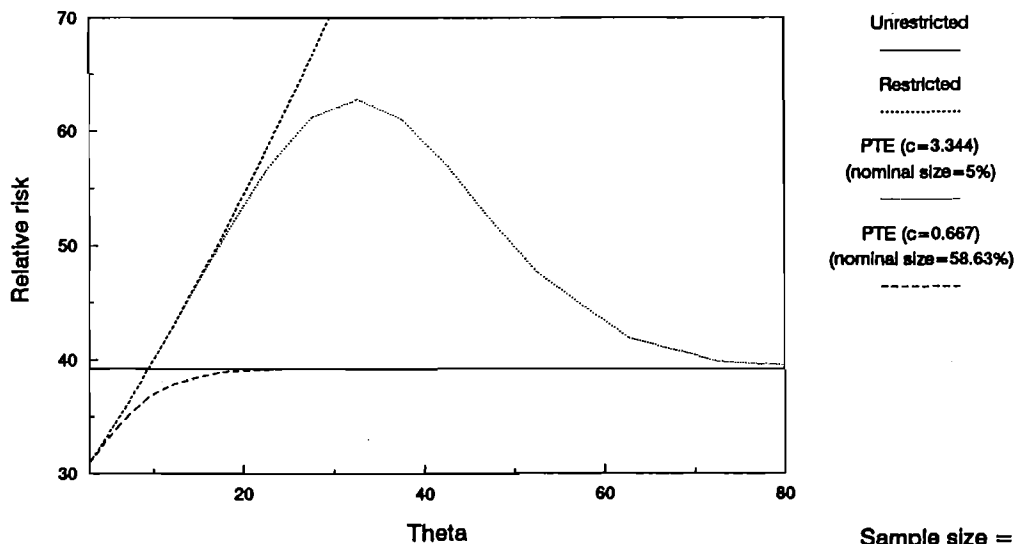


Application of Chow test
Excluded regressor is $N(0,1)$ r.v.

Sample size = 20
 $\lambda_2 = 2.5$
 $\theta_{\text{bar}} = 0.011$

FIGURE 7.6: Scale Parameter Estimator Risk Fns.

MS Estimators, Excluded Regressor, $\Psi = 1.0$
Regressors; Aus\$-US\$ x-rate & exponential trend



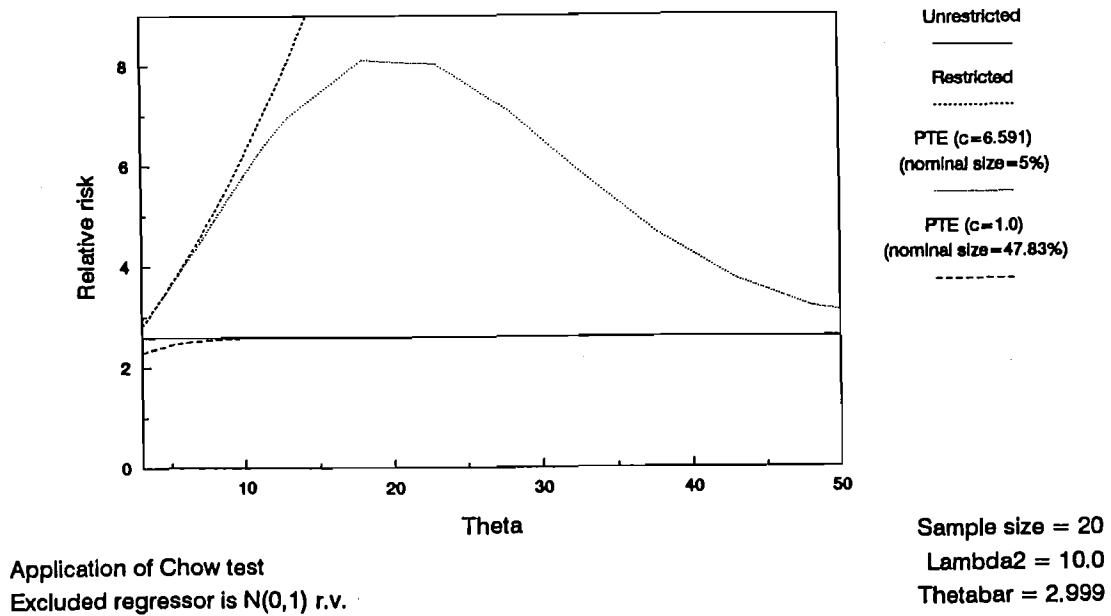
Application of Chow test
Excluded regressor is $N(0,1)$ r.v.

Sample size = 20
 $\Lambda_2 = 50.0$
 $\bar{\theta} = 2.755$

Note that, although in the absence of heteroscedasticity, these functions are independent of the particular data used, other than through θ , λ_2 and the degrees of freedom, the actual value of $\bar{\theta}$ will vary with the data. For example, while it is apparent that the restricted LS estimator has a lower risk than the unrestricted LS estimator over some part of the parameter space in Figure 7.4, this is not true in Figure 7.7 as the minimum value of θ , given this data set, is greater than that value at which $\rho(s_{LS}^*, \sigma^2) = \rho(s_{LS}, \sigma^2)$. A similar situation may arise in the case of the MS family of estimators. In general, however, the exclusion of relevant regressors appears unlikely to lead to a situation in which the restricted LS or MS estimators are dominated by their unrestricted counterparts. That this is not true in the case of the ML estimators is shown by Giles and Clarke (1989). Even if $\bar{\theta} \approx 0$ the restricted ML estimator may be strictly dominated by the unrestricted ML estimator.

FIGURE 7.7: Scale Parameter Estimator Risk Fns.

LS Estimators, Excluded Regressor, $\Psi = 1.0$
Regressors; Australian GDP & exponential trend



The effect of the exclusions on the pre-test estimators also varies depending on the particular family of regressors under consideration. In the case of the LS estimators, the PTE associated with a critical value of $c = 1$ dominates the unrestricted LS estimator despite the exclusions, as shown in Figures 7.4 and 7.7. Note, however, that the PTE associated with a higher critical value, $c = 6.531$, may be strictly dominated if $\bar{\theta}$ is sufficiently large. It is this PTE that would arise if the nominal size of the pre-test was chosen to be 5%. As shown by Giles and Clarke (1989), the ML PTE may be strictly dominated by the unrestricted ML estimator regardless of the critical value chosen for the pre-test, if the mis-specification caused by the exclusion of relevant regressors is severe. This is shown in Figures 7.2 and 7.5. In the case of MS estimators the exclusions appear to have little impact other than the distortion in the pre-test power function and the effect on $\bar{\theta}$ noted above. The PTE associated with a critical value of

$c = \nu/(\nu+2)$ dominates the unrestricted MS estimator as it does when there is no regressor mis-specification.

We now turn to a consideration of the effects of a mis-specified error covariance matrix on the estimators, given that the model is already mis-specified by the exclusion of relevant regressors. As in the case where there are no excluded regressors, quantitatively the risk of the estimators increases as ψ increases above unity, for a given value of θ and λ_2 . As ψ decreases below unity, the risk of the restricted estimators decreases while the risk of the unrestricted estimator may increase or decrease slightly depending on the data, the value of λ_2 and which of the LS, ML or MS family of estimators is being considered. The effect on the pre-test estimator as ψ decreases below unity depends also on the critical value chosen for the pre-test as we would expect, with relatively large critical values causing the pre-test estimator risk to lie close to the restricted estimator's risk, and therefore decreasing as ψ decreases; and relatively small critical values leading to a pre-test risk similar to the unrestricted estimator risk.

If the restricted estimator is not dominated by the unrestricted estimator, the range of θ over which the restricted estimator risk is lower than the unrestricted estimator risk increases as ψ decreases below unity and increases as ψ increases above unity. This is true regardless of the family of estimators being considered.

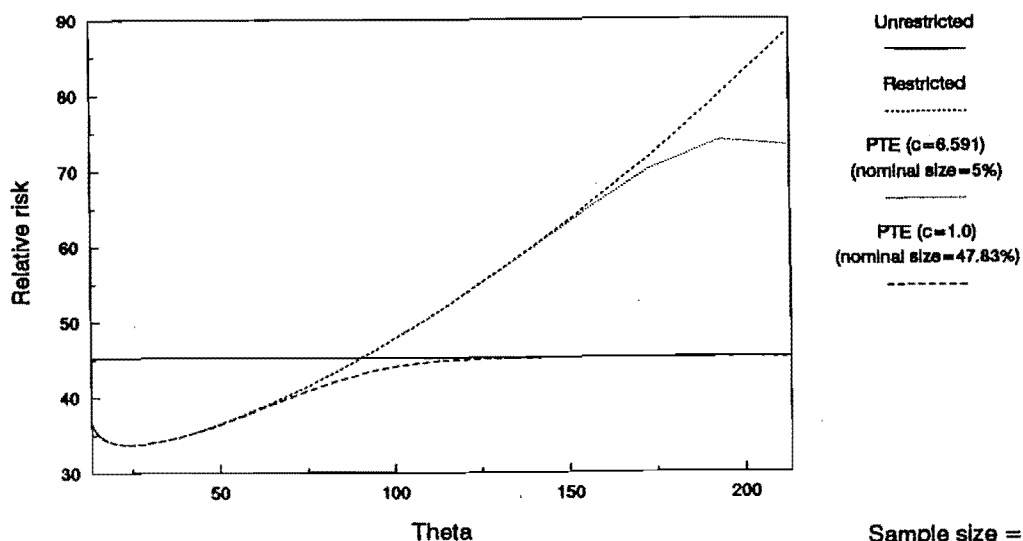
The effect of the heteroscedasticity on the pre-test estimator depends, as it does in the case where there are no excluded regressors, on which of the LS, ML or MS family of estimators is being considered. In general, however, the effects are qualitatively the same regardless of the particular regressors under consideration, other than the effect of the exclusion on $\bar{\theta}$ as noted above.

Recall that, when there is no regressor mis-specification, the pre-test LS estimator associated with a critical value of unity dominates the unrestricted LS estimator when ψ is greater than, or equal to, one. This is not necessarily so when ψ is less than unity. A similar result is apparent when there are regressors excluded from the model. For a given value of λ_2 , an increase in the value of ψ above unity has the same effect as such a change in the case where $\lambda_2 = 0$. That is, the PTE remains the dominating LS estimator when $c = 1$ and may also become more attractive relative to the restricted LD estimator. In contrast to the case where there are no excluded regressors, this effect will also occur as ψ decreases below unity if λ_2 is sufficiently large, as shown in Figure 7.8.

This occurs even if the restricted LS estimator and the pre-test LS estimator associated with a traditionally small nominal pre-test size are strictly dominated by the unrestricted LS estimator.

FIGURE 7.8: Scale Parameter Estimator Risk Fns.

LS Estimators, Excluded Regressor, Psi = 0.01
Regressors; New Zealand GDP & exponential trend

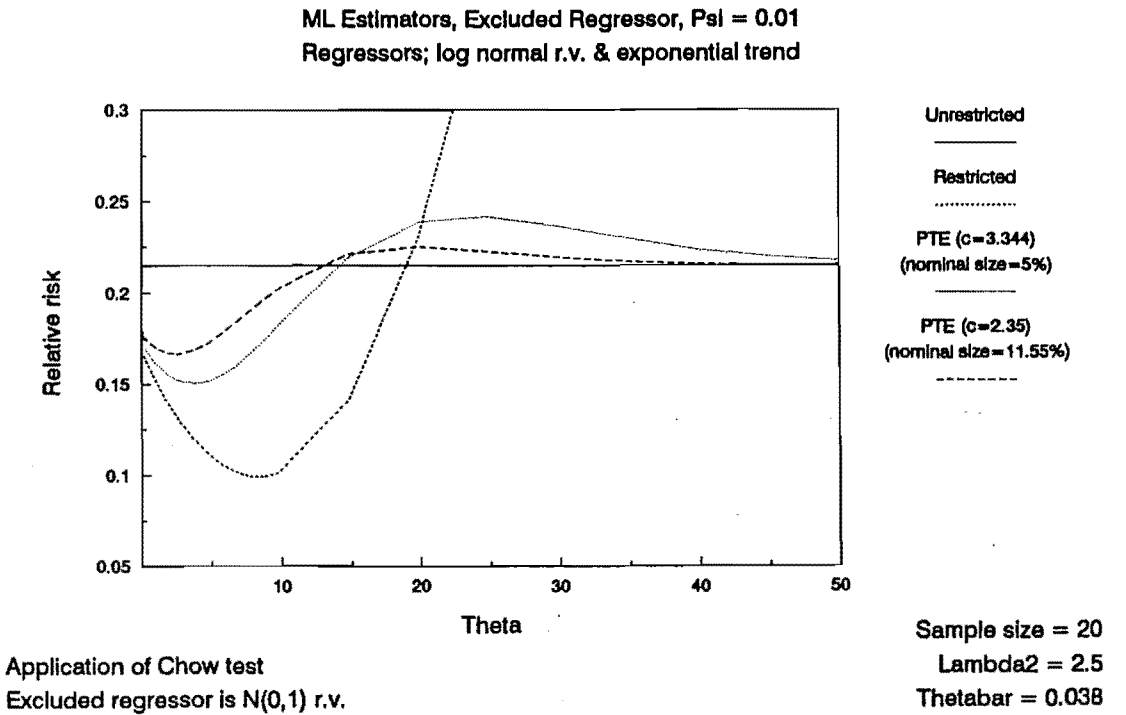


Application of Chow test
Excluded regressor is $N(0,1)$ r.v.

Sample size = 20
Lambda2 = 50.0
Thetabar = 13.192

The effect of the heteroscedasticity on the ML estimators depends on the degree of regressor mis-specification. Recall that, when there are no excluded regressors, the pre-test ML estimators and the restricted ML estimator may be dominated by the unrestricted ML estimator as ψ increases above unity, while they may become more attractive, relative to the unrestricted ML estimator as ψ decreases below unity. If there are regressors excluded from the model, similar effects are observed for moderate levels of λ_2 . For example, it may well be the case that the pre-test ML estimators are strictly dominated when there is no heteroscedasticity in the model, as in Figure 7.5. However, as ψ decreases below unity, the pre-test and restricted ML estimators may not be strictly dominated, as the effects of the heteroscedasticity offset (to some extent) the effects of the excluded regressors. This is shown in Figure 7.9.

FIGURE 7.9: Scale Parameter Estimator Risk Fns.



For larger values of λ_2 , however, there is little qualitative effect of the heteroscedasticity, and the unrestricted and pre-test ML estimators remain strictly dominated by the unrestricted ML estimator.

As in the case where there is no regressor mis-specification, for a given value of λ_2 , the effect of uncorrected heteroscedasticity on the MS family of estimators is qualitatively the same as the effect of heteroscedasticity on the LS family of estimators. For small values of λ_2 the pre-test MS estimator associated with a critical value of $c = \nu/(\nu+2)$ dominates the unrestricted MS estimator when ψ is equal to, or greater than, unity but it may not dominate when ψ takes a value less than unity as, Figure 7.10 illustrates. However, as the degree of regressor mis-specification increases, the PTE associated with $c = \nu/(\nu+2)$ becomes the dominating estimator regardless of the degree of heteroscedasticity in the model, as Figures 7.11 and 7.12 illustrate.

FIGURE 7.10: Scale Parameter Estimator Risk Fns.

MS Estimators, Excluded Regressor, Psi = 0.1
 Regressors; Aus\$-US\$ x-rate & exponential trend

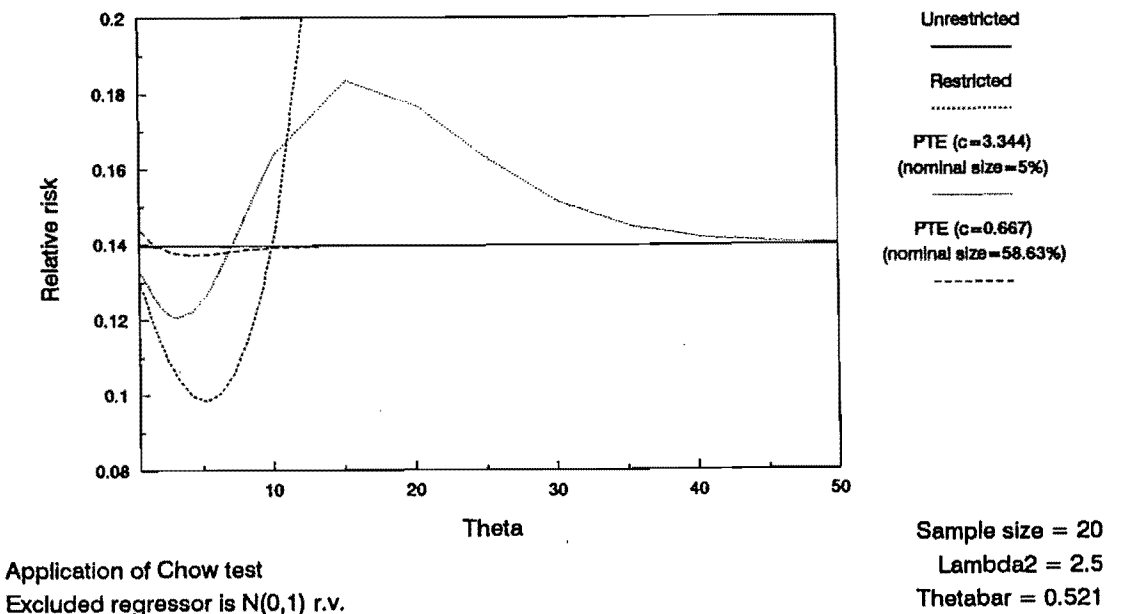
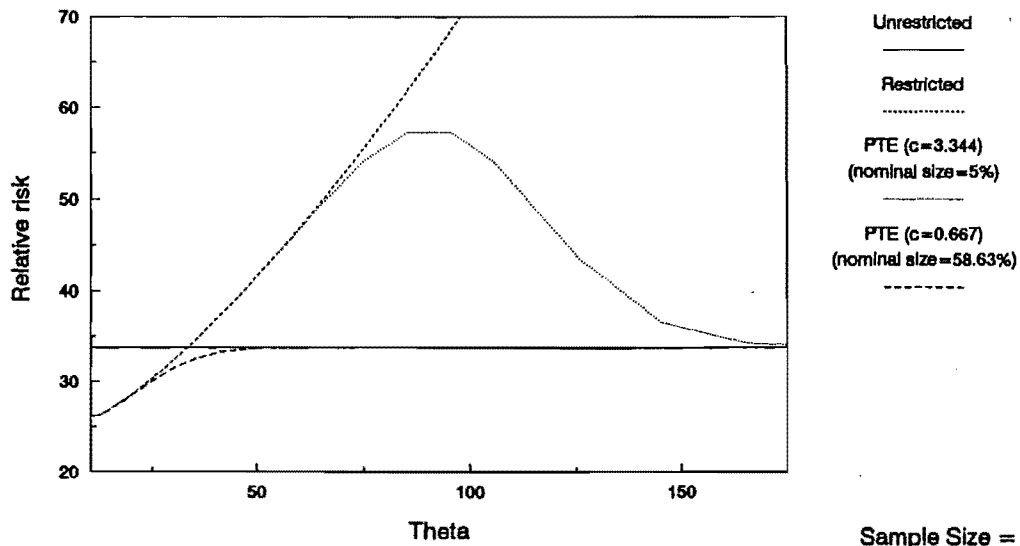


FIGURE 7.11: Scale Parameter Estimator Risk Fns.

MS Estimators, Excluded Regressor, $\Psi = 0.1$
Regressors; Aus\$-US\$ x-rate & exponential trend

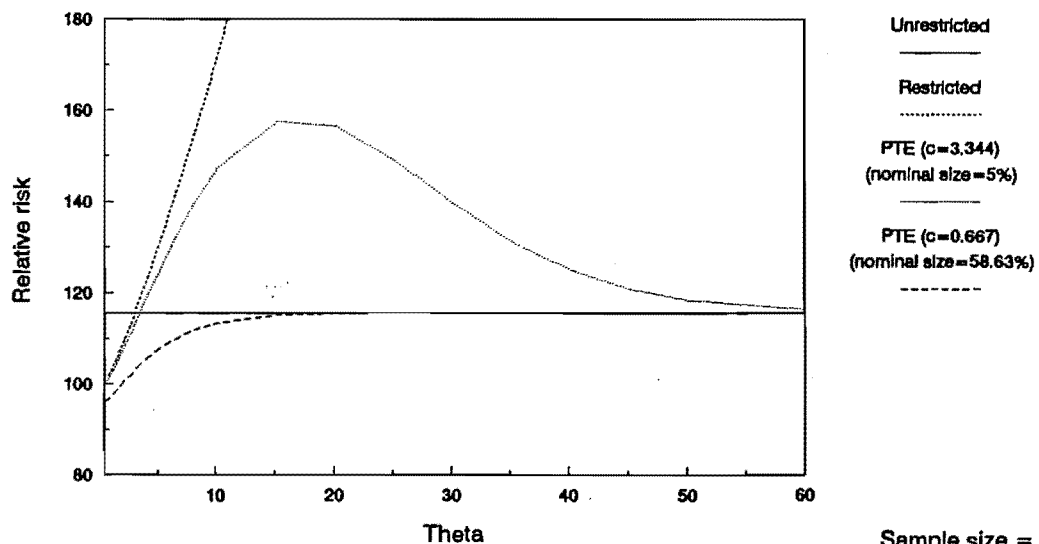


Application of Chow test
Excluded regressor is $N(0,1)$ r.v.

Sample Size = 20
 $\Lambda_2 = 50.0$
 $\bar{\theta} = 10.403$

FIGURE 7.12: Scale Parameter Estimator Risk Fns.

MS Estimators, Excluded Regressor, $\Psi = 10.0$
Regressors; Aus\$-US\$ x-rate & exponential trend



Application of Chow test
Excluded regressor is $N(0,1)$ r.v.

Sample size = 20
 $\Lambda_2 = 50.0$
 $\bar{\theta} = 0.465$

As is the case with the LS family of estimators, this occurs even in a situation where the restricted MS estimator and the pre-test estimators associated with a traditionally low nominal size are strictly dominated as a result of a large $\bar{\theta}$ value. To see why this is so consider Figure 7.12. If a different regressor had been excluded there would be little qualitative difference in the risk functions for each level of θ , though the value of $\bar{\theta}$ would change. Assume that the value of $\bar{\theta}$ increased to $\bar{\theta} = 10$. This would mean that, given the data and regressors, it was not possible for θ to be less than 10 in value. It is apparent that in this case the restricted MS estimator would be dominated over the entire parameter space. However, the PTE associated with a critical value of $c = \nu/(\nu+2)$ would not be dominated and would, in fact, still dominate the unrestricted estimator.

7.5 Conclusion

The practical implications of the double mis-specification of the model are similar to the implications under the situation discussed in Chapter Six, where there is only one mis-specification, namely that of the error covariance matrix. It is clear that the possibility of having excluded relevant regressors from the model further emphasizes the need to choose appropriate critical values for the pre-test rather than simply use critical values associated with nominal sizes of 1% or 5%. This is because in the presence of regressor mis-specification and uncorrected heteroscedasticity, the pre-test LS or MS estimators associated with critical values in the neighbourhoods of $c = 1$ and $c = \nu/(\nu+2)$ respectively appear likely to dominate their unrestricted counterparts.

If the ML estimators are to be used, either one of the two types of mis-specification considered may cause the pre-test ML estimator to be

strictly dominated and, although the effects of the heteroscedasticity may offset the effects of the excluded regressors to some extent if λ_2 is small, if λ_2 is greater than 7 in value the pre-test estimator is likely to be inadmissible. As in an applied situation the true value of λ_2 is unobservable, it would not be possible to determine whether or not the ML PTE is admissible for any particular mis-specified model.

Notwithstanding this, the ML estimators appear to be the most robust of the three families of estimators considered here, to uncorrected heteroscedasticity for a given value of λ_2 . Furthermore, it is apparent that, for a given level of ψ , the ML estimators also appear to be the most robust of the three families to increasing levels of regressor mis-specification. In an applied situation, therefore, if there is some doubt regarding the homoscedasticity of the errors, the unrestricted ML estimator appears to be the best estimator to apply.

APPENDIX 7A

Proof of Theorem 7.1.

$$\begin{aligned} s_i^2 &= (T+\Delta)^{-1}(y-Xb)'(y-Xb) \\ &= (T+\Delta)^{-1}(\varepsilon+\xi)'M(\varepsilon+\xi) \end{aligned}$$

hence a) and b) follow by Lemma 6.1 and Searle (1982, p.351,357).

$$\begin{aligned} s_i^{*2} &= (T+\Gamma)^{-1}(y-Xb^*)'(y-Xb^*) \\ &= (T+\Gamma)^{-1}M\bar{\varepsilon} + \bar{\varepsilon}'XCX'\bar{\varepsilon} \end{aligned}$$

where $\bar{\varepsilon} = (\varepsilon+X\eta\delta+\xi)$, $\bar{\varepsilon} \sim N(\bar{\delta}, \sigma^2\Omega)$, hence c) and d) follow by Lemma 6.1 and Searle (1982, p.351, 357). □

Proof of Corollary 7.1.

When $\Omega = I_T$

- a) $B(s_1^2, \sigma^2) = (T+\Delta)^{-1}(\text{tr}(M)+2\lambda_2) - 1 = (T+\Delta)^{-1}(T-K+2\lambda_2-(T+\Delta))$.
- b) $\begin{aligned} \rho(s_1^2, \sigma^2) &= (T+\Delta)^{-2} \left[(\text{tr}(M\Omega))^2 + 2\text{tr}(M)^2 + \frac{4}{\sigma^2} \xi' M \xi + \frac{2}{\sigma^2} \xi' M \xi \text{tr}(M) + \frac{1}{\sigma^4} (\xi' M \xi)^2 \right. \\ &\quad \left. - 2(T+\Delta)(\text{tr}M + \frac{1}{\sigma^2} \xi' M \xi) + (T+\Delta)^2 \right] \\ &= (T+\Delta)^{-2} \left[(T-K)^2 + 2(T-K) + 8\lambda_2 + 4\lambda_2(T-K) + 4\lambda_2^2 - 2(T+\Delta)(T-K) \right. \\ &\quad \left. - 4(T-K)\lambda_2 + (T+\Delta)^2 \right] \\ &= (T+\Delta)^{-2} (2(\nu+4\lambda_2) + (2\lambda_2 - K - \Delta)^2). \end{aligned}$
- c) $\begin{aligned} B(s_i^{*2}, \sigma^2) &= (T+\Gamma)^{-1} \left(\text{tr}(M+XCX') + \frac{1}{\sigma^2} (\delta' \eta' X' + \xi') XCX' (\xi + X\eta\delta) + \frac{1}{\sigma^2} \xi' M \xi - (T+\Gamma) \right) \\ &= (T+\Gamma)^{-1} (J+2\lambda_1+2\lambda_2-(K+\Gamma)). \end{aligned}$
- d) $\rho(s_i^{*2}, \sigma^2) = (T+\Gamma)^{-2} \left(\text{tr}(\phi) \right)^2 + 2\text{tr}(\phi)^2 + \frac{4}{\sigma^2} \bar{\delta}' \phi \bar{\delta}' + \frac{2}{\sigma^2} \bar{\delta}' \phi \bar{\delta} \text{tr}(\phi)$

$$\begin{aligned}
& + \frac{1}{\sigma^4} (\bar{\delta}' \phi \bar{\delta})^2 - 2(T+\Gamma) \left(\text{tr}(\phi) + \frac{1}{\sigma^2} \bar{\delta}' \phi \bar{\delta} \right) + (T+\Gamma)^2 \\
& = (T+\Gamma)^{-2} \left[(J+T-K)^2 + 2(J+T-K) + 8(\lambda_1 + \lambda_2) + 4(\lambda_1 + \lambda_2)(J+T-K) \right. \\
& \quad \left. + 4(\lambda_1 + \lambda_2)^2 - 2(T+\Gamma) \left((J+T-K) + 2(\lambda_1 + \lambda_2) \right) + (T+\Gamma)^2 \right] \\
& = (T+\Gamma)^{-2} \left[2(J+T+4(\lambda_1 + \lambda_2)) + (J-K-\Gamma+2(\lambda_1 + \lambda_2))^2 \right]. \quad \square
\end{aligned}$$

Proof of Theorem 7.2.

Given the appropriate definition of ζ , $\equiv \sigma^{-1} \Upsilon' \Omega^{-1/2} (X\eta\delta + \xi)$ in this case the proof is substantially the same as the proof of Theorems 6.2 and 6.3, except note that

$$E(z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z) = \sigma^2 \text{tr} M \Omega + \xi' M \xi$$

and

$$\begin{aligned}
E[(z' \Upsilon' \Omega^{1/2} M \Omega^{1/2} \Upsilon z)^2] & = \sigma^4 (\text{tr}(M \Omega))^2 + \sigma^2 \text{tr}(M \Omega)^2 + 4\sigma^2 \xi' M \Omega M \xi \\
& + 2\sigma^2 \xi' M \xi \text{tr}(M \Omega) + (\xi' M \xi)^2. \quad \square
\end{aligned}$$

Proof of Corollary 7.2.

Recall that the probability values, the P_{ij} 's, tend towards 0 (unity) as c tends towards 0 (∞). The proof of a) and b) is trivial. In the case of c) we have

$$\begin{aligned}
\lim_{c \rightarrow \infty} B(\hat{\sigma}_1^2, \sigma^2) & = (T+\Delta)^{-1} \left(\text{tr}(M \Omega) + \frac{1}{\sigma^2} \xi' M \xi \right) \\
& + (\Delta - \Gamma)(T+\Delta)^{-1} (T+\Gamma)^{-1} \text{tr} \left(M \Omega^{1/2} \Upsilon \Upsilon' (I_T + \zeta \zeta') \Upsilon' \Omega^{1/2} \right) \\
& + (T+\Gamma)^{-1} \text{tr} \left(X C X' \Omega^{1/2} \Upsilon (I_T + \zeta \zeta') \Upsilon' \Omega^{1/2} \right) \\
& = (T+\Gamma)^{-1} \left(\text{tr}(\phi \Omega) + \zeta' \Upsilon' \Omega^{1/2} \phi \Omega^{1/2} \Upsilon \zeta \right) - 1 \\
& = B(s_1^{*2}, \sigma^2)
\end{aligned}$$

as

$$\zeta' \Upsilon' \Omega^{1/2} \phi \Omega^{1/2} \Upsilon \zeta = \frac{1}{\sigma^2} (\delta' \eta' X' + \xi') \Omega^{-1/2} \Upsilon \Upsilon' \Omega^{1/2} (M + X C X') \Omega^{1/2} \Upsilon \Upsilon' \Omega^{-1/2} (\xi + X \eta \delta)$$

$$= \frac{1}{\sigma^2} \xi' M \xi + \frac{1}{\sigma^2} (\delta' \eta' X' + \xi') X C X' (X \eta \delta + \xi).$$

Similarly

$$\begin{aligned} \lim_{\sigma \rightarrow \infty} \rho(\hat{\sigma}_1^2, \sigma^2) &= ((T+\Delta)(T+\Gamma))^{-2} \left((\text{tr}(M\Omega))^2 + 2\text{tr}(M\Omega)^2 + \frac{4}{\sigma^2} \xi' M \Omega M \xi \right. \\ &\quad \left. + \frac{2}{\sigma^2} \xi' M \xi \text{tr}(M\Omega) + \frac{1}{\sigma^4} (\xi' M \xi)^2 \right) \\ &\quad + (2T+\Gamma+\Delta)(\Delta-\Gamma) \left(\frac{1}{\sigma^4} (\xi' M \xi)^2 + \frac{4}{\sigma^2} \xi' M \Omega M \xi + \frac{2}{\sigma^2} \xi' M \xi \text{tr}(M\Omega) \right. \\ &\quad \left. + 2\text{tr}(M\Omega)^2 + (\text{tr}(M\Omega))^2 \right) \\ &\quad + 2(T+\Delta)^2 \left(\frac{1}{\sigma^4} \bar{\delta}' M \bar{\delta} \bar{\delta} X C X' \bar{\delta} + \frac{4}{\sigma^2} \bar{\delta}' M \Omega X C X' \bar{\delta} \right. \\ &\quad \left. + \frac{1}{\sigma^2} \xi' M \xi \text{tr}(X C X' \Omega) + \frac{1}{\sigma^2} \bar{\delta}' X C X' \bar{\delta} \text{tr}(M\Omega) \right. \\ &\quad \left. + \text{tr}(X C X' \Omega) \text{tr}(M\Omega) + 2\text{tr}(X C X' \Omega M \Omega) \right) \\ &\quad + (T+\Delta)^2 \left(\frac{1}{\sigma^4} (\bar{\delta}' X C \bar{X}' \bar{\delta})^2 + \frac{4}{\sigma^2} \bar{\delta}' X C X' \Omega X C X' \bar{\delta} + \frac{2}{\sigma^2} \bar{\delta}' X C X' \bar{\delta} \text{tr}(X C X' \Omega) \right. \\ &\quad \left. + 2\text{tr}(X C X' \Omega)^2 + (\text{tr}(X C X' \Omega))^2 \right) \\ &\quad - \frac{2}{T+\Gamma} \left(\text{tr}(\phi\Omega) + \frac{1}{\sigma^2} \bar{\delta}' \phi \bar{\delta} \right) + 1 \\ &= \rho(s_1^{*2}, \sigma^2). \end{aligned}$$

□

Proof of Corollary 7.3.

Recall that, when $\Omega = I_T$, Υ can be chosen as $\bar{\Upsilon}$, the joint eigenvector matrix of $X C X'$, $X S^{-1} X'$, M and $(X C X' - \frac{cJ}{T-K} M)$ and also that

$$\lambda_i = \sigma^2 \text{ for } i = 1, \dots, J,$$

$$\lambda_i = 0 \text{ for } i = J+1, \dots, K$$

and

$$\lambda_i = -\frac{cJ}{T-K} \sigma^2 \text{ for } i = K+1, \dots, T.$$

by Lemma 4.1.

Hence, when $\Omega = I_T$,

$$\begin{aligned} B(\hat{\sigma}_1^2, \sigma^2) &= (T+\Delta)^{-1} \left(\text{tr}(M\Omega) + \frac{1}{\sigma^2} \xi' M \xi \right) + (\Delta - \Gamma)(T+\Delta)^{-1} (T+\Gamma)^{-1} \Gamma' M \Gamma B + \frac{1}{T+\Gamma} \Gamma' X C X' \Gamma B - 1 \\ &= ((T+\Delta)(T+\Gamma))^{-1} \left[(T+\Gamma)^2 (\nu + 2\lambda_2) + (\Delta - \Gamma) (\nu h'(0,2) + 2\lambda_2 h'(0,4)) \right. \\ &\quad \left. + (T+\Delta) (J h'(2,0) + 2\lambda_1 h'(4,0)) \right] - 1 \end{aligned}$$

and

$$\begin{aligned} \rho(\hat{\sigma}_1^2, \sigma^2) &= ((T+\Delta)(T+\Gamma))^{-2} \left[(T+\Gamma)^2 (\nu^2 + 2\nu + 8\lambda_2 + 4\lambda_2 \nu + 4\lambda_2^2) \right. \\ &\quad + (2T+\Gamma+\Delta)(\Delta - \Gamma) (4\lambda_2^2 h'(0,8) + \nu(\nu+2)h'(0,4) + 4\lambda_2(\nu+2)h(0,6)) \\ &\quad + 2(T+\Delta)^2 + (4\lambda_1 \lambda_2 h'(4,4) + J\nu h'(2,2) + 2J\lambda_2 h'(2,4) \\ &\quad + 2\nu\lambda_1 h'(4,2)) + (T+\Delta)^2 (4\lambda_1^2 h'(8,0) + J(J+2)h'(4,0) + 4\lambda_1(J+2)h'(6,0)) \\ &\quad - 2(T+\Gamma)^2 (T+\Delta)(\nu + 2\lambda_2) - 2(T+\Gamma)(T+\Delta)(\Delta - \Gamma) (\nu h'(0,2) + 2\lambda_2 h'(0,4)) \\ &\quad \left. - 2(T+\Gamma)(T+\Delta)^2 (J h'(2,0) + 2\lambda_1 h'(4,0)) + (T+\Delta)^2 (T+\Gamma)^2 \right] \\ &= (T+\Delta)^{-2} (2(\nu + 4\lambda_2) + (2\lambda_2 - K - \Delta)^2) \\ &\quad + ((T+\Delta)(T+\Gamma))^{-2} \left[(T+\Delta)^2 \left(2J\nu h'(2,2) - 2(T+\Gamma)J h'(2,0) + 4\nu\lambda_1 h'(4,2) \right. \right. \\ &\quad + (J(J+2) - 4(T+\Gamma)\lambda_1) h'(4,0) + 4\lambda_1(J+2)h'(6,0) \\ &\quad \left. \left. + 4\lambda_1^2 h'(8,0) + 4\lambda_2 (J h'(2,4) + 2\lambda_1 h'(4,4)) \right) \right. \\ &\quad + (\Delta - \Gamma)(2T+\Delta+\Gamma) \left(4\lambda_2 (\lambda_2 h'(0,8) + (\nu+2)h'(0,6)) \right. \\ &\quad \left. \left. + \nu(\nu+2)h'(0,4) \right) - 2(T+\Gamma)(T+\Delta)(\Delta - \Gamma) (\nu h'(0,2) + 2\lambda_2 h'(0,4)) \right] \quad \square \end{aligned}$$

CHAPTER EIGHT

THE ROBUSTNESS OF OPTIMAL PRE-TEST CRITICAL VALUES IN A LINEAR MODEL WITH A MIS-SPECIFIED ERROR COVARIANCE MATRIX

8.1 Introduction

As the properties of the linear restrictions PTE depend, in part, on the critical value chosen for the pre-test, a number of authors have considered the problem of determining the optimal pre-test critical value or size under a given optimality criterion. These authors have focussed primarily on optimal critical values when estimating the prediction vector or, equivalently, estimating the coefficient vector in a model with orthonormal regressors, as the estimator risk functions in this case are independent of the data. Two such authors are, for example, Toyoda and Wallace (1976) and Brook (1976). Both Brook and Toyoda and Wallace propose rules of thumb for the determination of the optimal critical value for pre-testing linear restrictions in the linear model. These rules of thumb are approximately invariant to increases in the sample size and number of regressors in the model.

More recently Brook and Fletcher (1981) suggest a similar rule of thumb for determining an appropriate critical value when estimating the coefficient vector in a model with some degree of collinearity between the regressors. It has, however, been shown that (uncorrected) autocorrelation or heteroscedasticity in the regression disturbance can distort the pre-test power function (see, for example, Schmidt and Sickles (1977), Kiviet (1980), Consiglieri (1981), Giles and Scott (1992), Giles and Lieberman (1991) and Chapter Four of this thesis, among others). The question therefore arises of the robustness of these rules of thumb to such a mis-specification of the

error covariance matrix. It is this issue that we address here. Following the approach of Brook and Fletcher (1981), we determine the optimal critical values for a pre-test of linear restrictions when estimating the coefficient vector in a number of regression models with a mis-specified error covariance matrices.

In section 8.2 we discuss the differing optimality criteria that have been considered in earlier papers and their benefits and drawbacks. Exact optimal critical values are determined for a number of regression models and are reported and discussed in section 8.3. Section 8.4 concludes the chapter.

8.2 Optimality Criteria

Consider the classical linear regression model

$$y = X\beta + \varepsilon ; \varepsilon \sim N(0, \sigma^2 \Omega)$$

where y , X , β and ε are as defined above. In addition to the sample information we will assume that there is non-sample information which is expressed in the form of the null hypothesis; $H_0: R\beta = r$ vs $H_A: R\beta \neq r$. This hypothesis is tested using the usual Wald statistic

$$u = \frac{(Rb-r)'(RS^{-1}R')^{-1}(Rb-r)T-K}{(y-Xb)'(y-Xb) J} ,$$

where b is the OLS estimator of β and $S = X'X$. If the null hypothesis is rejected, the OLS estimator, b , is applied to the model to estimate β , otherwise the RLS estimator, b^* , is applied. This procedure gives rise to the classical PTE

$$\hat{\beta} = \begin{cases} b = S^{-1}X'y & ; u \geq c \\ b^* = b + S^{-1}R'(RS^{-1}R')^{-1}(r-Rb) & ; u < c \end{cases}$$

where c is the critical value chosen for the pre-test. The properties of the PTE therefore depend in part on the value of c and it is the problem of

choosing a value of c such that the PTE risk is optimised under some appropriate criterion which we consider here.

The risks of the OLS, RLS and PT estimators in this situation are given in Chapter Four of this thesis and are

$$\rho(b, \beta) = \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1}),$$

$$\rho(b^*, \beta) = \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1}) - 2S^{-1}X' \Omega X C + C X' \Omega X C + \delta' \eta' \eta \delta$$

and

$$\begin{aligned} \rho(\hat{\beta}, \beta) &= \sigma^2 \text{tr}(S^{-1}X' \Omega X S^{-1}) - 2\sigma^2 \text{tr}(C X' \Omega^{1/2} \Gamma B \Gamma' \Omega^{1/2} X S^{-1}) \\ &\quad + 2\delta' \eta' C X' \Omega^{1/2} \Gamma P_3 \Gamma' \Omega^{-1/2} X \eta \delta \\ &\quad + \sigma^2 \text{tr}(C X' \Omega^{1/2} \Gamma B \Gamma' \Omega^{1/2} X C) \end{aligned}$$

respectively, where $\eta = S^{-1}R'(RS^{-1}R')^{-1}$, $C = S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$, Γ is the $T \times T$ eigenvector matrix of $\Phi \equiv \Omega^{1/2}(XCX' - \frac{cJ}{T-K}(I_T - XS^{-1}X'))\Omega^{1/2}$, and P_3 and B are as defined above in Chapter Four.

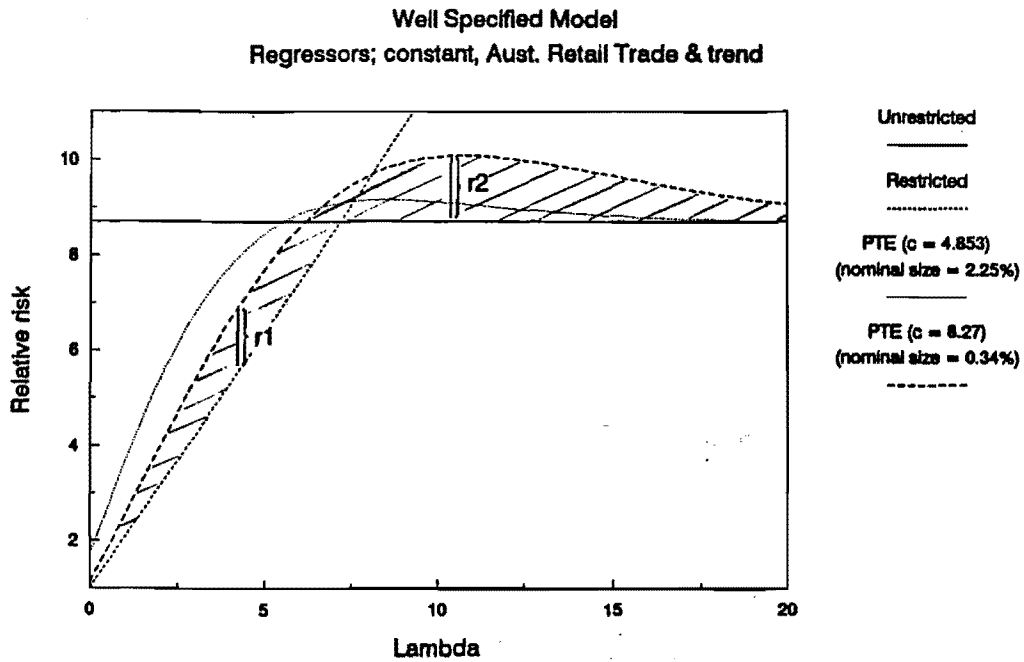
This then is an extension to the problem of determining optimal critical values in a well specified model, i.e. one in which $\Omega = I_T$, with non-orthonormal regressors considered by Brook and Fletcher (1981). The two optimality criteria considered by Brook and Fletcher are the minimum average risk criterion of Toyoda and Wallace (1976) and the minimax regret criterion of Brook (1976). These criteria can be illustrated using Figure 8.1 which shows OLS, RLS and PT estimator risk functions for the case of estimating the coefficient vector in a well specified model.

Let the value of the non-centrality parameter, $\lambda = \frac{1}{2\sigma^2} \delta'(RS^{-1}R')^{-1} \delta$,

where $\delta = R\beta - r$, at which the RLS risk is equal to the OLS risk be denoted λ^* . When $\lambda < \lambda^*$ the minimum risk estimator (of the three under consideration) is the RLS estimator, b^* , otherwise the minimum risk estimator is the OLS estimator, b .

The minimum average risk criterion of Toyoda and Wallace (1976) consists of choosing a value of c such that the area between the PTE risk function and the minimum risk boundary is minimized. That is, it involves of choosing a value of c such that the shaded area in Figure 8.1 is minimized.

FIGURE 8.1: Coefficient Estimator Risk Functions



The minimax regret criteria adopted by Brook (1976), among others, consists of choosing a value of c such that the regret, defined for a given value of λ as $\text{Regret}(c) = \rho(\hat{\beta}(c), \beta) - \min_{\bar{c}} \left[\rho(\hat{\beta}(\bar{c}), \beta) \right]$, is minimized, where $\hat{\beta}(\bar{c})$ denotes the PTE associated with a critical value of \bar{c} ; $\bar{c} \in [0, \infty)$. Note, however, that the minimum PTE risk equals the RLS risk for $\lambda \leq \lambda^*$ and equals the OLS risk for $\lambda > \lambda^*$. Hence the problem becomes $\min_c \left[\max_{\lambda} \left[\rho(\hat{\beta}(c), \beta) - \min(\rho(b, \beta), \rho(b^*, \beta)) \right] \right]$. The optimal critical value will therefore be one that minimizes the greater of the two regrets r_1 and r_2 in Figure 8.1. In fact, this implies that r_1 and r_2 will be equal in magnitude at the optimal choice of c .

In this chapter we consider finding the optimal critical value for the pre-test in a mis-specified model under the criterion¹ of $\min_c \left[\max_{\theta} \left(\rho(\hat{\beta}(c), \beta) - \min(\rho(b, \beta), \rho(b^*, \beta)) \right) \right]$ where $\theta = \sum_{(i: \lambda_i > 0)}^1 (v_i' \Omega^{-1/2} X \eta \delta)^2$, λ_i is the i 'th eigenvalue, and v_i the i 'th eigenvector, of Φ . We choose this criterion for two reasons; firstly because it has received a good deal of attention in the literature (see Sawa and Hiromatsu (1971), Gun (1967), Brook (1976), and Giles and Lieberman (1991b) for example), and secondly because the robustness of optimal critical values to various mis-specifications has been considered using this criterion by Giles, Lieberman and Giles (1992), who consider the robustness of Brook's (1976) optimal critical values to regressor mis-specification, and Wong and Giles (1991), who consider the problem when the error distribution is mis-specified as normal when it is, in fact, spherically symmetric.

In determining the optimal pre-test critical values, Brook and Fletcher (1981) scale the regressor matrix so that the $X'X$ matrix is in correlation form² to afford some protection from the fact that the estimators' risks are not scale invariant. In their worked example Brook and Fletcher consider the model of Marquardt and Snee (1975) which involves a full quadratic polynomial in each of three predictor variables. In correlation form $X'X$ matrix is a 9×9 matrix, implying that the dependent variable is also expressed in deviations about means and hence there is no constant term in the regressor

¹ This criterion is equivalent to Brook's (1976) criterion when the model is well specified. However, when the model is mis-specified it is possible for the PTE to be the minimum risk estimator over some part of the parameter space. Wong and Giles (1991) use this criterion in a similar situation in estimating optimal critical values in a linear spherically symmetric errors.

² This is achieved by expressing the regressors as deviations about the mean and scaling so that $x_i' x_i = 1$, where x_i is the i 'th regressor.

set. Therefore the risk functions used by Brook and Fletcher represent the risk of the coefficient vector excluding the coefficient on the constant. Although we follow Brook and Fletcher in scaling $X'X$ to correlation form, for the sake of consistency with the main body of the literature and earlier chapters of this thesis, we will continue to include a constant term in our regressor set. This will not lessen the degree of correspondence between our results and Brook and Fletcher's results as the optimal critical values will be the same regardless of this.

The calculation of the optimal critical values has been carried out for a number of different linear models and data sets on a VAX 6340 using a FORTRAN program written by the author and incorporating Davies' (1980) algorithm. We use an iterative process to determine the maximum values of regret, denoted r_1 and r_2 in Figure 8.1 for each value of the critical value, c , considered. Each step in this process requires T^2 calls on Davies' (1980) algorithm, itself an iterative procedure. Having determined r_1 and r_2 for the value of c under consideration, we proceed to iterate c towards the optimal value and repeat our calculation of r_1 and r_2 with this new value of c .

This process is, therefore, computationally burdensome and the number of specific cases we consider is limited as a result. In addition, the effects of the mis-specification are data dependent and hence any optimal critical value we calculate would be useful in an applied situation only if the researcher had full knowledge of the mis-specification, in which case the model would presumably be estimated in such a way as to correct for the problem.

For these reasons it is not our intention to provide exhaustive tables, which would be of limited use in a practical situation in any case, but rather to consider the robustness of Brook's (1976) and Brook and Fletcher's

(1981) commonly applied rules of thumb to a mis-specification of the error covariance matrix.

8.3 Numerical Results

The particular models we consider in calculating optimal critical values are those described in Chapter Four. Although the error term in these models is generated by one of a number of autoregressive or heteroscedastic processes, we will assume that this is not taken into account when estimating the models. With the exception of the mis-specification of the error covariance matrix, the usual assumptions of the classical linear model hold and the regressor matrix is well specified. The data we use to evaluate these models are described in Appendix 4B. Recall that from Chapter Four, an uncorrected positive autoregressive process in the errors is likely to cause an upwards bias in the pre-test's power function. Intuitively we would expect such a bias to result in an increase in the optimal critical value of the pre-test. The more highly trended the regressors being tested, the greater the distortion in power, and hence the greater the expected increase in the optimal critical value c^* . As the distortion is more pronounced in the case of AR(1) errors, as compared with the other autoregressive or heteroscedastic processes that we consider, it follows that the increase in the value of c^* will also be more pronounced in this case.

The numerical results support these conjectures, as is apparent from Table 8.1. Note that the regressor matrix is orthonormal in this case and hence the optimal critical value in a well-specified model is $c^* = 1.89$, the value calculated by Brook (1976) for this case. The true optimal level of significance, denoted "true α^* " in Table 8.1, remains fairly constant at around 15 to 20% regardless of the degree and type of mis-specification of the model. The actual critical values, however, do vary. They generally

increase with ρ in the case of AR(1) and MA(1) errors, while no clear pattern is apparent in the case of AR(4) and heteroscedastic errors. The effect of the heteroscedasticity is not as great as the effect of autoregressive errors and again this is consistent with our earlier observations regarding the degree of distortion in the pre-test power function.

A roughly similar pattern is evident in Tables 8.2 and 8.4, although in some cases, for example with the Australian Retail trade regressor, the effect of the additional trend variable is to cause the optimal critical value to decrease with increasing values of ρ . In these tables, as in Table 8.1, the optimal critical value in a well-specified model is calculated by Brook (1976) to be $c^* = 1.89$.

In Table 8.3 we see the effect of an uncorrected AR(4) error process on the optimal pre-test critical values when testing the significance of a group of seasonal dummy variables. As any increase in ρ introduces a large upwards distortion in the pre-test's power function, it is hardly surprising that the optimal critical values also increase with ρ .

Note that, in Tables 8.3 and 8.4, the critical values suggested by Brook and Fletcher's rule of thumb are not necessarily good approximations for the actual calculated optimal critical value in a well-specified model. Consider the cases of testing the joint significance of Australian Retail Trade and a linear trend, in Table 8.4, or testing the joint significance of seasonal dummies in Table 8.3. These are illustrated in Figures 8.1 and 8.2 respectively. Brook and Fletcher's rule understates, in the former case, and overstates, in the latter case, the actual calculated optimals. As Brook and Fletcher do not justify their rule of thumb, it is not possible to determine why it is not as accurate in these cases as it is in others.

TABLE 8.1: Optimal Pre-Test Critical Values and Sizes

T = 18, K = 2, J = 1

Data	c [*]	true α	nompl. α	c [*]	true α	nompl. α	c [*]	true α	nompl. α	c [*]	true α	nompl. α	c [*]	true α	nompl. α
Autoregressive Errors	$\rho=-0.9$			$\rho=-0.6$			$\rho=0.0$			$\rho=0.6$			$\rho=0.9$		
AR(1) Errors															
Normal r.v. with AR(1)	0.49	0.242	0.491	0.96	0.194	0.342	1.89	0.189	0.189	3.33	0.188	0.088	4.24	0.185	0.057
Australian CPI	0.28	0.214	0.603	0.56	0.206	0.467	1.89	0.189	0.189	7.08	0.205	0.018	24.13	0.214	0.000
Australian Retail Trade	2.54	0.158	0.146	1.42	0.180	0.301	1.89	0.189	0.189	5.81	0.190	0.029	16.08	0.180	0.001
Australian Trade Balance	1.08	0.176	0.315	0.98	0.185	0.338	1.89	0.189	0.189	3.99	0.193	0.064	6.04	0.183	0.027
AR(4) Errors															
Australian CPI	0.71	0.187	0.413	1.00	0.199	0.333	1.89	0.189	0.189	2.26	0.199	0.154	1.06	0.216	0.319
Australian Retail Trade	0.32	0.206	0.581	0.78	0.196	0.391	1.89	0.189	0.189	2.82	0.206	0.114	2.92	0.209	0.108
Australian Trade Balance	4.57	0.208	0.050	3.09	0.194	0.099	1.89	0.189	0.189	1.69	0.191	0.213	1.61	0.202	0.224
Uniformly Dist. r.v.	2.92	0.201	0.108	2.49	0.202	0.135	1.89	0.189	0.189	1.12	0.196	0.308	0.78	0.201	0.391
MA(1) Errors															
Normal r.v. with AR(1)	1.06	0.195	0.320	1.15	0.193	0.300	1.89	0.189	0.189	2.70	0.195	0.121	2.80	0.197	0.115
Australian CPI	0.28	0.195	0.604	0.45	0.193	0.514	1.89	0.189	0.189	3.59	0.197	0.078	3.85	0.199	0.069
Australian Retail Trade	0.80	0.182	0.386	0.87	0.193	0.365	1.89	0.189	0.189	2.97	0.197	0.105	3.15	0.198	0.096
Australian Trade Balance	0.73	0.189	0.406	0.84	0.192	0.374	1.89	0.189	0.189	3.19	0.188	0.094	3.21	0.200	0.093
Heteroscedastic Errors	$\psi=0.625$			$\psi=0.25$			$\psi=1.0$			$\psi=4.0$			$\psi=16.0$		
Model (b)															
Australian GDP	1.94	0.189	0.182	2.04	0.193	0.172	1.89	0.189	0.189	1.74	0.193	0.205	1.68	0.198	0.213
Uniformly Dist. r.v.	1.97	0.189	0.180	2.12	0.190	0.165	1.89	0.189	0.189	1.68	0.189	0.213	1.59	0.189	0.225
Australian\$ to U.S.\$ rate	1.84	0.189	0.194	1.77	0.193	0.202	1.89	0.189	0.189	2.01	0.194	0.175	2.07	0.200	0.169
Model (c)															
Australian GDP	2.03	0.189	0.173	2.45	0.189	0.137	1.89	0.189	0.189	1.71	0.189	0.210	1.78	0.190	0.200
Uniformly Dist. r.v.	2.07	0.189	0.170	2.55	0.189	0.130	1.89	0.189	0.189	1.56	0.189	0.230	1.45	0.190	0.246
Australian\$ to U.S.\$ rate	1.99	0.189	0.172	2.26	0.191	0.152	1.89	0.189	0.189	1.73	0.191	0.208	1.73	0.195	0.208
Model (d)															
Australian GDP	1.76	0.189	0.203	1.88	0.191	0.193	1.89	0.189	0.189	1.78	0.214	0.200	2.38	0.189	0.143
Uniformly Dist. r.v.	2.03	0.189	0.173	2.40	0.190	0.141	1.89	0.189	0.189	1.65	0.189	0.218	1.64	0.190	0.219
Australian\$ to U.S.\$ rate	1.89	0.189	0.188	2.03	0.188	0.174	1.89	0.189	0.189	2.15	0.188	0.162	2.78	0.187	0.115
Model (e)															
Australian GDP	2.04	0.189	0.172	2.51	0.188	0.133	1.89	0.189	0.189	1.65	0.189	0.217	1.64	0.189	0.218
Uniformly Dist. r.v.	1.90	0.189	0.187	1.91	0.189	0.186	1.89	0.189	0.189	1.80	0.189	0.198	1.69	0.190	0.212
Australian\$ to U.S.\$ rate	2.01	0.189	0.175	2.41	0.187	0.14	1.89	0.189	0.189	1.75	0.188	0.205	1.83	0.189	0.195
Model (f)															
Australian GDP	1.97	0.189	0.180	2.23	0.189	0.155	1.89	0.189	0.189	1.84	0.189	0.194	2.01	0.191	0.166
Uniformly Dist. r.v.	1.94	0.189	0.183	2.00	0.190	0.176	1.89	0.189	0.189	1.75	0.183	0.204	1.50	0.183	0.204
Australian\$ to U.S.\$ rate	1.99	0.188	0.177	2.10	0.189	0.167	1.89	0.189	0.189	1.78	0.214	0.201	2.35	0.190	0.145

TABLE 8.2: Optimal Pre-Test Critical Values and Sizes

T = 18, K = 3, J = 1

Regressors are Data Vector and Linear Trend

Data	c [*]	trye _α	nompl. _α	c [*]	trye _α	nompl. _α	c [*]	trye _α	nompl. _α	c [*]	trye _α	nompl. _α	c [*]	trye _α	nompl. _α
Autoregressive Errors	$\rho=-0.9$			$\rho=-0.6$			$\rho=0.0$			$\rho=0.6$			$\rho=0.9$		
AR(1) Errors															
Normal r.v. with AR(1)	0.67	0.208	0.426	1.08	0.191	0.314	1.80	0.199	0.199	2.52	0.208	0.132	2.73	0.246	0.118
Australian CPI	0.84	0.204	0.373	1.12	0.188	0.307	1.87	0.189	0.189	3.61	0.189	0.076	5.75	0.178	0.029
Australian Retail Trade	7.86	0.154	0.013	3.29	0.180	0.088	1.80	0.199	0.199	1.70	0.195	0.210	1.35	0.235	0.262
Australian Trade Balance	1.67	0.168	0.214	1.28	0.179	0.275	1.80	0.199	0.199	2.76	0.206	0.116	2.62	0.229	0.125
MA(1) Errors															
Normal r.v. with AR(1)	1.25	0.200	0.281	1.35	0.192	0.262	1.80	0.199	0.199	2.48	0.199	0.135	2.56	0.202	0.129
Australian CPI	1.06	0.199	0.319	1.18	0.191	0.294	1.87	0.189	0.189	2.75	0.194	0.117	2.85	0.198	0.111
Australian Retail Trade	2.13	0.189	0.164	2.10	0.190	0.167	1.80	0.199	0.199	1.75	0.173	0.204	1.57	0.192	0.228
Australian Trade Balance	1.13	0.181	0.303	1.20	0.184	0.290	1.80	0.199	0.199	2.69	0.202	0.121	2.84	0.202	0.111
Heteroscedastic Errors	$\psi=0.625$			$\psi=0.25$			$\psi=1.0$			$\psi=4.0$			$\psi=16.0$		
Model (b)															
Australian GDP	2.00	0.191	0.177	2.18	0.195	0.159	1.89	0.189	0.189	1.62	0.187	0.221	1.50	0.186	0.239
Uniformly Dist. r.v.	2.00	0.189	0.177	2.18	0.190	0.159	1.89	0.189	0.189	1.63	0.191	0.220	1.51	0.192	0.237
Australian\$ to U.S.\$ rate	1.93	0.189	0.184	2.00	0.189	0.177	1.89	0.189	0.189	1.78	0.195	0.200	1.74	0.200	0.206
Model (c)															
Australian GDP	1.94	0.191	0.183	2.14	0.195	0.163	1.89	0.189	0.189	1.86	0.187	0.192	1.97	0.185	0.180
Uniformly Dist. r.v.	2.03	0.191	0.174	2.40	0.195	0.141	1.89	0.189	0.189	1.65	0.187	0.218	1.58	0.186	0.227
Australian\$ to U.S.\$ rate	1.90	0.191	0.187	1.98	0.195	0.178	1.89	0.189	0.189	1.96	0.187	0.181	2.11	0.187	0.165
Model (d)															
Australian GDP	1.93	0.191	0.184	2.09	0.195	0.168	1.89	0.189	0.189	1.90	0.186	0.188	2.05	0.184	0.172
Uniformly Dist. r.v.	1.99	0.191	0.177	2.27	0.195	0.152	1.89	0.189	0.189	1.74	0.187	0.206	1.79	0.187	0.200
Australian\$ to U.S.\$ rate	1.90	0.190	0.187	1.91	0.191	0.186	1.89	0.189	0.189	1.79	0.193	0.200	1.64	0.193	0.218
Model (e)															
Australian GDP	1.90	0.200	0.187	2.23	0.194	0.155	1.89	0.189	0.189	1.81	0.187	0.197	1.88	0.185	0.190
Uniformly Dist. r.v.	1.90	0.190	0.187	1.89	0.191	0.188	1.89	0.189	0.189	1.73	0.198	0.207	1.61	0.193	0.223
Australian\$ to U.S.\$ rate	1.00	0.331	0.332	1.01	0.300	0.330	1.89	0.189	0.189	1.92	0.189	0.185	1.90	0.189	0.187

TABLE 8.3: Optimal Pre-Test Critical Values and Sizes

T = 21, K = 5, J = 3

Regressors are Data Vector and Seasonal Dummies

Data	c	true α	nompl. α	c	true α	nompl. α	c	true α	nompl. α	c	true α	nompl. α	c	true α	nompl. α	Brook and Fletcher's suggested c
Autoregressive Errors	$\rho=-0.9$			$\rho=-0.6$			$\rho=0.0$			$\rho=0.6$			$\rho=0.9$			
AR(4) Errors																
Australian CPI	0.21	0.283	0.888	0.25	0.459	0.861	1.37	0.288	0.288	6.37	0.298	0.005	34.84	0.311	0.000	2.55
Australian Retail Trade	0.20	0.299	0.893	0.40	0.302	0.754	1.23	0.330	0.330	5.62	0.329	0.008	28.65	0.356	0.000	2.86
Australian Trade Balance	0.58	0.181	0.639	0.62	0.237	0.613	1.39	0.283	0.283	5.52	0.306	0.009	28.51	0.319	0.000	2.64
Uniformly Dist. r.v.	0.21	0.257	0.891	0.25	0.515	0.861	1.39	0.282	0.282	6.36	0.261	0.005	29.25	0.319	0.000	2.62

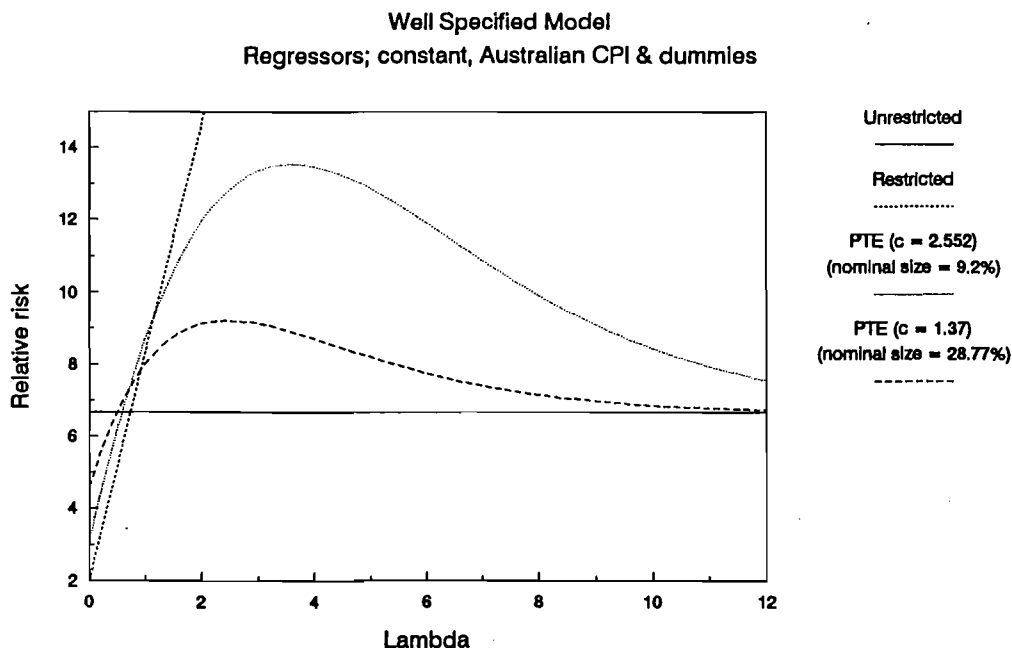
TABLE 8.4: Optimal Pre-Test Critical Values and Sizes

T = 19, K = 3, J = 2

Regressors are Data Vector and Linear Trend

Data	c	true α	nompl. α	c	true α	nompl. α	c	true α	nompl. α	c	true α	nompl. α	c	true α	nompl. α	Brook and Fletcher's suggested c
Autoregressive Errors	$\rho=-0.9$			$\rho=-0.6$			$\rho=0.0$			$\rho=0.6$			$\rho=0.9$			
AR(1) Errors																
Normal r.v. with AR(1)	0.39	0.255	0.687	0.74	0.219	0.494	1.60	0.233	0.233	4.24	0.239	0.033	12.94	0.237	0.000	2.17
Australian CPI	245.31	0.000	0.000	304.45	0.000	0.000	339.25	0.000	0.000	963.90	0.000	0.000	*****	0.000	0.000	242.06
Australian Retail Trade	53.63	0.000	0.000	13.31	0.000	0.000	8.27	0.003	0.003	9.45	0.056	0.002	23.00	0.105	0.000	4.85
Australian Trade Balance	1.81	0.016	0.195	1.42	0.077	0.270	2.49	0.114	0.114	6.67	0.134	0.008	20.24	0.139	0.000	2.18
AR(4) Errors																
Australian CPI	593.87	0.000	0.000	543.68	0.000	0.000	339.25	0.000	0.000	475.94	0.000	0.000	413.43	0.000	0.000	242.06
Australian Retail Trade	3.36	0.001	0.060	5.17	0.002	0.019	8.27	0.003	0.003	17.22	0.002	0.000	26.06	0.000	0.000	4.85
Australian Trade Balance	5.20	0.097	0.018	3.30	0.096	0.063	2.49	0.114	0.114	2.99	0.105	0.079	3.17	0.066	0.069	2.18
Uniformly Dist. r.v.	2.55	0.147	0.113	2.09	0.157	0.156	2.12	0.166	0.166	2.05	0.167	0.161	1.39	0.172	0.276	2.00
MA(1) Errors																
Normal r.v. with AR(1)	0.73	0.199	0.496	0.83	0.203	0.454	1.60	0.233	0.233	2.65	0.239	0.101	2.80	0.242	0.091	2.17
Australian CPI	286.88	0.000	0.000	309.29	0.000	0.000	339.25	0.000	0.000	717.81	0.000	0.000	747.85	0.000	0.000	242.06
Australian Retail Trade	8.74	0.001	0.003	8.65	0.001	0.003	8.27	0.003	0.003	8.17	0.016	0.004	8.15	0.019	0.004	4.85
Australian Trade Balance	1.20	0.070	0.328	1.33	0.076	0.293	2.49	0.114	0.114	4.24	0.125	0.033	4.49	0.128	0.028	2.18

FIGURE 8.2: Coefficient Estimator Risk Functions



Note, however, that the critical values suggested by Brook and Fletcher's rule are closer to the optimal critical values than those values associated with traditional pre-test sizes of 5% and 1%. Note also that we are constrained by the computational cost associated with evaluating some of the risk functions to considering examples with relatively small sample sizes.

The effect of an increase in the sample size on the optimal critical value is illustrated in Table 8.5. As in this table we consider only one restriction, the optimal critical values calculated by Brook (1976) apply for the case in which the model is well-specified. The effect of the mis-specification on the estimators' risk functions varies with the data and we have seen, in Chapter Four, examples in which the restricted estimator are strictly dominated by the unrestricted estimator at one sample size, although it is not if we choose a different sample or a different sized sample for the regression.

TABLE 8.5: The Effects of a Change in Sample Size

	c [*]	true α	nompl. α	c [*]	true α	nompl. α	c [*]	true α	nompl. α	c [*]	true α	nompl. α	c [*]	true α	nompl. α
Sample Size	ρ=-0.9			ρ=-0.6			ρ=0.0			ρ=0.6			ρ=0.9		
Regressors; constant and Australian Trade Balance. Testing the significance of Australian Trade Balance.															
AR(1) Errors															
T = 6	1.51	0.213	0.287	1.92	0.193	0.238	1.92	0.238	0.238	0.00	1.000	1.000	0.00	1.000	1.000
T = 10	0.56	0.095	0.475	0.74	0.217	0.415	1.90	0.206	0.206	6.27	0.267	0.037	11.58	0.262	0.009
T = 18	1.12	0.114	0.306	0.99	0.255	0.336	1.89	0.189	0.189	3.98	0.271	0.063	5.10	0.157	0.038
T = 24	1.39	0.128	0.249	0.92	0.263	0.347	1.88	0.183	0.183	2.27	0.268	0.145	0.00	1.000	1.000
AR(4) Errors															
T = 6	0.67	0.650	0.459	1.06	0.371	0.361	1.92	0.238	0.238	2.38	0.199	0.198	2.60	0.189	0.182
T = 10	2.44	0.235	0.157	2.37	0.259	0.162	1.90	0.206	0.206	1.10	0.265	0.325	0.41	0.203	0.538
T = 18	4.31	0.237	0.054	3.14	0.273	0.095	1.89	0.189	0.189	1.55	0.260	0.231	1.45	0.202	0.247
T = 24	2.68	0.245	0.226	3.04	0.239	0.095	1.88	0.183	0.183	0.00	1.000	1.000	0.00	1.000	1.000
Regressors; constant, Australian Trade Balance, linear trend and seasonal dummies. Testing the significance of linear trend.															
AR(1) Errors															
T = 10	0.82	0.214	0.417	1.04	0.222	0.365	1.92	0.238	0.238	3.83	0.253	0.122	7.94	0.256	0.048
T = 14	0.73	0.137	0.418	0.82	0.183	0.392	1.90	0.206	0.206	6.69	0.210	0.032	19.30	0.214	0.002
T = 22	0.73	0.172	0.405	0.70	0.170	0.415	1.89	0.189	0.189	7.36	0.196	0.015	27.91	0.206	0.000
T = 30	0.72	0.047	0.404	0.67	0.153	0.422	1.88	0.183	0.183	7.00	0.202	0.014	28.57	0.222	0.000
AR(4) Errors															
T = 10	0.53	0.293	0.507	1.05	0.254	0.364	1.92	0.238	0.238	2.70	0.235	0.176	3.06	0.235	0.155
T = 14	0.59	0.442	0.463	1.02	0.282	0.342	1.90	0.206	0.206	3.02	0.194	0.120	3.65	0.195	0.092
T = 22	0.00	1.000	1.000	0.67	0.264	0.425	1.89	0.189	0.189	4.20	0.186	0.057	5.93	0.191	0.027
T = 30	0.50	0.163	0.486	0.88	0.163	0.357	1.88	0.183	0.183	4.53	0.197	0.044	7.24	0.205	0.013
Regressors; constant, exponential trend, linear trend and seasonal dummies. Testing the significance of exponential trend.															
AR(1) Errors															
T = 10	1.01	0.228	0.372	1.01	0.236	0.371	1.92	0.238	0.238	4.62	0.240	0.098	6.81	0.239	0.059
T = 14	0.73	0.221	0.418	0.84	0.209	0.386	1.90	0.206	0.206	4.87	0.232	0.058	7.58	0.264	0.025
T = 22	0.28	0.369	0.603	0.67	0.206	0.424	1.89	0.189	0.189	5.48	0.217	0.033	8.28	0.323	0.011
T = 30	0.00	1.000	1.000	0.60	0.210	0.447	1.88	0.183	0.183	5.99	0.199	0.022	13.17	0.257	0.001
Regressors; constant, Australian Retail Trade, linear trend and seasonal dummies. Testing the significance of Australian Retail Trade.															
AR(1) Errors															
T = 10	1.63	0.129	0.271	1.37	0.180	0.307	1.92	0.238	0.238	1.39	0.522	0.304	0.00	1.000	1.000
T = 14	0.71	0.232	0.423	0.89	0.212	0.374	1.90	0.206	0.206	4.65	0.228	0.063	4.40	0.397	0.069
T = 22	0.93	0.202	0.349	0.98	0.194	0.336	1.89	0.189	0.189	3.54	0.197	0.078	3.14	0.288	0.095
T = 30	1.41	0.175	0.247	1.13	0.183	0.299	1.88	0.183	0.183	3.29	0.188	0.082	2.97	0.254	0.098

Under our minimax regret criterion the optimal critical value, in a situation in which the restricted estimator is strictly dominated, is $c^* = 0$, that is, we should always apply the unrestricted estimator.

This value of $c^* = 0$ appears in a number of cells in Table 8.5 implying that, in these particular cases, the restricted estimator is dominated. With the exception of these cells, the general pattern is that the optimal values generally increase with ρ , and the true optimal sizes do not vary greatly as a result of the mis-specification.

8.4 Conclusion

It is apparent that the exact optimal critical values based on a mini-max regret criterion are not robust to a mis-specification of the error covariance matrix. We have seen that the optimal critical values generally increase with ρ in the case of an AR(1) or MA(1) process in the errors and that the optimal true size of the pre-test is largely unaffected by the mis-specification. However, there are situations in which the optimal critical value is $c^* = 0$ with an associated optimal size of $\alpha^* = 100\%$. Without knowing the form of the true error covariance matrix it is not possible to know in an applied situation if this situation will arise.

Because the exact critical values are not robust to a mis-specification of this type, it follows that the rules of thumb proposed to approximate them will not be robust either (even if such rules are accurate, which may not be the case). Hence the use of an "optimal" critical value, based on such a rule, will not necessarily lead to an optimal risk. This finding is analogous to the findings of Giles, Lieberman and Giles (1992) who consider the determination of optimal critical values in a model mis-specified by the exclusion of relevant regressors, and Wong and Giles (1991) who consider the problem in the context of a model with Multivariate Student-t disturbances.

CHAPTER NINE

CONCLUDING REMARKS

In this thesis we address the problem of pre-testing in a linear regression model that has been mis-specified as a result of incorrectly assuming a scalar error covariance matrix, and possibly also excluding relevant regressors from the model. We consider estimation of the coefficient vector, the prediction vector and the regression scale parameter in this framework, and derive the bias and risk (under quadratic loss) of the linear restrictions pre-test estimators of these parameters in this situation.

As the risk formulae of these estimators are complex, we numerically evaluate them for a number of linear models and data sets to illustrate the effect of the mis-specification(s) on the sampling properties of the estimators. In Chapter Four we consider the problem of estimating the coefficient vector in a model mis-specified as a result of incorrectly assuming a scalar error covariance matrix with no excluded relevant regressors. Our results show that the presence of (uncorrected) heteroscedasticity, or a positive autoregressive process in the errors, is likely to mean that the pre-test power function is distorted upwards, particularly if the regressors are trended and the errors are generated by an AR(1) or MA(1) process.

In addition, the unexpected presence of a non-scalar error covariance matrix may cause the restricted and pre-test estimators to be strictly dominated by the unrestricted estimator, in which case it is always best to apply the unrestricted estimator in terms of minimizing risk. In general we

find that the effect of an uncorrected AR(1) or AR(4) error process on estimator risks is greater than the effect of an uncorrected MA(1) or heteroscedastic error process, other things being equal.

It has been argued that autocorrelation may arise in the regression disturbance term as a result of the incorrect exclusion of relevant regressors from the design matrix. Therefore the problem of excluded regressors in conjunction with the problem of a non-scalar error covariance matrix naturally arises. This is the situation which we address in Chapter Five. In addition to considering the unrestricted, restricted and pre-test estimators of the coefficient vector, we also consider the corresponding estimators of the prediction vector in this context. Although the properties of the restricted, unrestricted and pre-test predictors in a model with excluded relevant regressors are well known, the implications of such a mis-specification, in terms of estimating the coefficient vector, have not, heretofore, been examined.

The pre-test itself will be distorted as a result of excluding relevant regressors from the model even if the error covariance matrix is well specified. As the power of the pre-test depends on λ_1 (the numerator non-centrality parameter of the pre-test statistic's distribution), while the risk functions of the restricted and pre-test estimators of the coefficient vector do not depend directly on λ_1 , it is possible that this pre-test estimator may be strictly dominated by the unrestricted estimator. This arises because the region of the λ_1 space in which the pre-test is likely to accept the restrictions may be a region in which the restricted estimator has a greater risk than the unrestricted estimator.

This situation does not occur if we consider the predictor estimators, as the restricted and pre-test predictors rely directly on λ_1 when the error

covariance matrix is well-specified and hence the region of the λ_1 space in which the restrictions are most likely to be accepted is also the region in which the restricted estimator attains its minimum risk. In particular, the restricted predictor will have a lower risk than the unrestricted predictor if $\lambda_1 < \frac{J}{2}$, where J is the number of linear restrictions being tested. (See Mittlehammer (1984)). This does not mean, however, that the restricted and pre-test estimators always have lower risk than the unrestricted estimator somewhere on the parameter space. We show that the minimum value of the parameter λ_1 may differ from 0 to the extent that both the unrestricted and pre-test estimators may be strictly dominated by the unrestricted estimator. The minimum value of λ_1 , which we may denote $\bar{\lambda}$, increases with λ_2 , the denominator non-centrality parameter of the distribution of the pre-test, and with the number of restrictions being tested, and generally decreases as the correlation between the included and excluded regressors increases. If, however, the pre-test is of the significance of those regressors which are correlated with the excluded regressors, $\bar{\lambda}$ increases with the level of this correlation. This implies that the included regressors may be acting as proxy variables for those that are excluded if they are highly correlated.

If the model is mis-specified as a result of incorrectly assuming a scalar error covariance matrix in addition to being mis-specified due to the exclusion of relevant regressors, the same general conclusions still apply. In fact, if the mis-specification of the regressor matrix is severe enough, the additional mis-specification of the error covariance matrix may make little quantitative difference to the estimators' risks, other than through the pre-test power function.

Frequently in an applied situation it is desired to estimate the scale parameter in addition to the coefficient or prediction vector. The

properties of the unrestricted, restricted and pre-test maximum likelihood (ML) estimators of the scale parameter in a model mis-specified by the incorrect assumption of a scalar covariance matrix, along with the corresponding least squares (LS) and minimum mean squared error (MS) estimators, are considered in Chapter Six where we derive formulae for their relative bias and risk. Because the general formulae are rather complex and computationally burdensome to evaluate, we consider a particular model in our numerical evaluations, in which the formulae simplify considerably.

Unlike the estimators of the coefficient vector, the estimators of the scale parameter have relative risk functions that are independent of the data, other than through the degrees of freedom and non-centrality parameter of the pre-test if the model is well specified. Because of this the effects of the mis-specification on a particular family of estimators are qualitatively the same regardless of which data set is considered. As in the case of estimating the coefficient vector, (uncorrected) heteroscedasticity, in addition to distorting the pre-test power function, may cause the restricted and pre-test estimators to dominate the unrestricted estimator over a greater or less part of the parameter space. In fact, it is possible for the ML restricted and pre-test estimators to be strictly dominated by the unrestricted estimator, regardless of the choice of pre-test critical value, c .

Recall that, in the case of the LS and MS family of estimators, there exists a family of pre-test estimators which strictly dominate their unrestricted counterparts (see Ohtani (1988) and Giles (1991a)). The presence of heteroscedasticity in the model may accentuate or cancel out this property of the pre-test estimators, although no case was found in which these families of pre-test LS or MS estimators were themselves strictly dominated. Despite the fact that the heteroscedasticity may cause the

pre-test ML estimator to become strictly dominated by the unrestricted ML estimator, it appears to be the most robust to this mis-specification of the three families of pre-test estimators considered, in terms of having the lowest risk in a mis-specified model. Given that the unrestricted ML estimator may well be risk superior to the corresponding PTE, it appears to be the best estimator to apply in circumstances where the usual assumption of a scalar error covariance is in doubt.

A similar observation can be made in the case considered in Chapter Seven, where we wish to estimate the scale parameter in a model possibly mis-specified in terms of having relevant regressors excluded from the regressor matrix in addition to having a non-scalar error covariance matrix. The effects of the exclusion of relevant regressors on the scale parameter estimators are similar to the effects of such an exclusion on the coefficient and predictor estimators. In the case of the ML estimators the excluded regressors may have the effect of causing the restricted and PT estimators to be dominated, even if $\bar{\lambda}$ is close to 0, this does not happen in the case of the LS and MS estimators. However, it may be the case that $\bar{\lambda}$ is sufficiently large to cause the restricted LS or MS estimators to be strictly dominated by the corresponding unrestricted estimators. Nonetheless, with a suitable choice of critical value, there exist pre-test estimators which dominate these unrestricted estimators.

If the mis-specification of the design matrix is severe enough, the additional effects of a mis-specification of the error covariance matrix are qualitatively minor, other than the indirect effect through the pre-test power function. Moreover it is apparent that, for the cases we consider, the unrestricted ML estimator is the most robust of those considered to double mis-specifications of this kind.

The distortion of the pre-test power function caused by a mis-specification of the error covariance matrix is likely to have an impact of the "rules of thumb" proposed in the literature by Brook (1976), Toyoda and Wallace (1976) and Brook and Fletcher (1981) to determine the optimal critical value for the pre-test. To examine these effects we consider in Chapter Eight a number of linear regression models and evaluate the exact optimal critical value for the pre-test in each one using the minimax regret approach of Brook (1976). It is apparent that the major impact of the mis-specification of the error covariance matrix, in terms of determining the optimal critical value arises from the distortion of the pre-test power function. Generally, the true optimal pre-test size is in the range of 17%-22%, regardless of the degree of model mis-specification, and the optimal critical value adjusts to maintain the optimal size. However, in some cases the restricted estimator may be dominated by the unrestricted estimator, and the optimal critical value falls to 0.

We have assumed throughout this thesis that no test for a non-scalar error covariance matrix has been carried out prior to estimation. A logical extension to this work would be to consider a pre-test for homoscedasticity in a linear model where there is autocorrelation in the errors as well as (possible) heteroscedasticity. This, then, would be an extension of the pooling problem considered by Bancroft (1944) and Toyoda and Wallace (1975), among others, in terms of estimating the error variance after such a preliminary test, and Greenberg (1980) and Mandy (1984), among others, who consider the related problem of estimating the coefficient vector in this situation.

A related question would be the one of estimating the coefficient vector and/or the scale parameter in a linear model after a preliminary test for

autocorrelation, where there is also some degree of heteroscedasticity present in the regression disturbances. Small (1991) and Giles and Small (1991) have recently provided some evidence regarding the properties of traditional pre-tests for autocorrelation in such circumstances. In addition, the problem of pre-testing using an autocorrelation or heteroscedasticity consistent covariance matrix, such as that suggested by Newey and West (1987) or White (1980), for example, has yet to be considered.

Despite the research that remains to be done on related topics, however, it is apparent that true properties of the commonly applied linear restrictions pre-test estimator may vary markedly from the assumed properties when no correction is made for a non-scalar error covariance matrix and/or there are relevant regressors excluded from the model. The exact nature of the distortion cannot be determined without knowledge of the particular mis-specification involved and the particular estimator that is used. Nonetheless, it is clear that potentially there are a number of hazards involved in mis-specifying the regression error process in this way.

REFERENCES

- Aigner, D.J. (1974), MSE dominance of least squares with errors-of-observation, *Journal of Econometrics* 2, 365-372.
- Aitken, A.C. (1934), On least squares and linear combinations of observations, *Proceedings of the Royal Society of Edinburgh* 55, 42-48.
- Amemiya, T. (1977), A note on a heteroscedastic model, *Journal of Econometrics* 6, 365-370.
- Balestra, P. (1980), A note on the exact transformation associated with first-order moving average process, *Journal of Econometrics* 14, 381-394.
- Bancroft, T.A. (1944), On biases in estimation due to the use of preliminary tests of significance, *Annals of Mathematical Statistics* 15, 190-204.
- Bancroft, T.A. and C-P. Han (1977), Inference based on conditional specification: A note and a bibliography, *International Statistical Review* 45, 117-127.
- Bartlett, M.S. (1937), Properties of sufficiency and statistical tests, *Proceedings of the Royal Society A* 160, 268-282.
- Battese, G.E. and B.P. Bonyhady (1981), Estimation of household expenditure functions: An application of a class of heteroscedastic regression models, *Economic Record* 57, 80-85.
- Berenblut, I.I. and G.I. Webb (1973), A new test for autocorrelated errors in the linear regression model, *Journal of the Royal Statistical Society B* 35, 33-50.
- Bock, M.E., T.A. Yancey and G.G. Judge (1973), The statistical consequences of the preliminary test estimators in regression, *Journal of the American Statistical Association* 68, 109-116.
- Box, G.E.P. (1952), Multi-factor designs of first order, *Biometrika* 39, 49-57.
- Breusch, T.S. and A.R. Pagan (1979), A simple test for heteroscedasticity and random coefficient variation, *Econometrica* 47, 1287-1294.
- Brook, R.J. (1972), On the use of minimax regret functions to set significance points in prior tests of estimation, Ph.D. thesis, North Carolina State University.
- Brook, R.J. (1976), On the use of a regret function to set significance points in prior tests of estimation, *Journal of the American Statistical Association* 71, 126-131.
- Brook, R.J. and R.H. Fletcher (1981), Optimal significance levels of prior tests in the presence of multicollinearity, *Communications in Statistics: Theory and Methods A* 10, 1401-1413.

- Chow, G.C. (1960), Tests of equality between sets of coefficients in two linear regressions, *Econometrica* 28, 591-605.
- Clarke, J.A. (1986), Some implications of estimating a regression scale parameter after a preliminary test of restrictions, M.Ec. thesis, Monash University.
- Clarke, J.A. (1990), Preliminary-test estimation of the standard error of estimate in linear regression, *Economics Letters* 34, 27-32.
- Clarke, J.A., D.E.A. Giles and T.D. Wallace (1987a), Estimating the error variance in regression after a preliminary test of restrictions on the coefficients, *Journal of Econometrics* 34, 293-304.
- Clarke, J.A., D.E.A. Giles and T.D. Wallace (1987b), Preliminary-test estimation of the error variance in linear regression, *Econometric Theory* 3, 299-304.
- Cohen, A. (1965), Estimates of the linear combinations of parameters in the mean vector of a multivariate distribution, *Annals of Mathematical Statistics* 36, 78-87.
- Consiglieri, I. (1981), The Chow test with serially correlated errors, *Rivista Internazionale di Scienze Sociali* 89, 125-137.
- Corsi, P., E. Pollack and S.L. Pracken (1982), The Chow test in the presence of serially correlated errors, in G.C. Chow and P. Corsi (eds.) *Evaluating the Reliability of Macroeconomic Models*, New York, McGraw-Hill.
- David, F.N. and J. Neyman (1938), Extension of the Markoff theorem in least squares, *Statistical Research Memoirs* 2, 105-116.
- Davies, R.B. (1980), The distribution of a linear combination of χ^2 random variables (Algorithm AS 155), *Applied Statistics* 29, 323-333.
- Durbin, J. and G.S. Watson (1950), Testing for serial correlation in least squares regression I, *Biometrika* 37, 409-428.
- Durbin, J. and G.S. Watson (1951), Testing for serial correlation in least squares regression II, *Biometrika* 38, 159-178.
- Durbin, J. (1970a), Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, *Econometrica* 38, 410, 421.
- Durbin, J. (1970b), An alternative to the bounds test for testing for serial correlation in least squares regression, *Econometrica* 38, 422-429.
- Dyer, D.D. and J.P. Keating (1980), On the determination of critical values for Barlett's test, *Journal of the American Statistical Association* 75, 313-319.
- Engle, R.F., D.F. Hendry and D. Trumble (1985), Small sample properties of ARCH estimators and tests, *Canadian Journal of Economics* 51, 66-93.

- Fomby T.B. and D.K. Guilkey (1978), On Choosing the optimal level of significance for the Durbin-Watson test and the Bayesian alternative, *Journal of Econometrics* 8, 203-213.
- Froehlich, B.R. (1973), Some estimators for a random coefficient regression model, *Journal of the American Statistical Association* 68, 329-334.
- Geary, R.C. (1966), A note on residual heterovariance and estimation efficiency in regression, *American Statistician* 20(4), 30-31.
- Gelfand, A.E. and D.K. Dey (1988), Improved estimation of the disturbance variance in a linear regression model, *Journal of Econometrics* 39, 387-395.
- Giles, D.E.A. (1986), Preliminary-test estimation in mis-specified regressions, *Economics Letters* 21, 325-328.
- Giles, D.E.A. (1992a), The exact distribution of a simple pre-test estimator, in W.E. Griffiths, H. Lütkepohl and M.E. Bock (eds.) *Readings in Econometric Theory and Practice: In Honor of George Judge*, Amsterdam, North Holland, 57-74.
- Giles, D.E.A. and M. Beattie (1987), Autocorrelation pre-test estimation in models with a lagged dependent variable, in M.L. King and D.E.A. Giles (eds.), *Specification Analysis in the Linear Model*, London, Routledge and Kegan Paul, 99-116.
- Giles, D.E.A. and J. Clarke (1989), Preliminary-test estimation of the scale parameter in a mis-specified regression model, *Economics Letters* 30, 201-205.
- Giles, D.E.A. and M.L. King (1978), Fourth order autocorrelation: Further significance points for the Wallis test, *Journal of Econometrics* 8, 255-259.
- Giles, D.E.A. and O. Lieberman (1991a), Bounds on the effects of heteroscedasticity on the Chow test for structural change, Discussion Paper 9101, Department of Economics, University of Canterbury.
- Giles, J.A. and O. Lieberman (1991b), The optimal size of a preliminary test for linear restrictions when estimating the regression scale parameter, *Economics Letters* 37, 25-30.
- Giles, D.E.A. & O. Lieberman (1992a), Some Properties of the Durbin-Watson test after a preliminary t-test, *Journal of Statistical Computation and Simulation* 41, 219-227.
- Giles, D.E.A. and O. Lieberman (1992b), Preliminary test estimation under absolute error loss. Mimeo, Department of Economics, University of Canterbury.
- Giles, D.E.A., O. Lieberman and J.A. Giles (1992), The optimal size of a preliminary test of linear restrictions in a mis-specified regression model, *Journal of the American Statistical Association*, 87, 1153-1157.

- Giles, D.E.A. and M. Scott (1992), Some consequences of using the Chow test in context of autocorrelated disturbances, *Economics Letters* 38, 145-150.
- Giles, D.E.A. and V.K. Srivastava (1993), The exact distribution of a least squares regression coefficient estimator after a preliminary t test, *Statistics and Probability Letters* 16, forthcoming.
- Giles, J.A. (1991a), Pre-testing for linear restrictions in a regression model with spherically symmetric disturbances, *Journal of Econometrics* 50, 377-398.
- Giles, J.A. (1991b), Pre-testing in a mis-specified regression model, *Communications in Statistics: Theory and Methods A* 20, 3221-3238.
- Giles, J.A. (1992b), Estimation of the error variance after a preliminary-test of homogeneity in a regression model with spherically symmetric disturbances, *Journal of Econometrics* 53, 345-361.
- Giles, J.A. and D.E.A. Giles (1991), Preliminary-test estimation of the regression scale parameter when the loss function is asymmetric, Discussion Paper 9104, Department of Economics, University of Canterbury.
- Giles, J.A. and D.E.A. Giles (1993), Pre-test estimation and testing in econometrics: Recent developments, *Journal of Economic Surveys*, forthcoming.
- Goldfeld, S.M. and R.E. Quandt (1965), Some tests for homoscedasticity, *Journal of the American Statistical Association* 60, 539-547.
- Gourieroux, C. and A. Trognan (1984), Specification pre-test estimator, *Journal of Econometrics* 25, 15-27.
- Greenberg, E. (1980), Finite sample moments of a preliminary-test estimator in the case of possible heteroscedasticity, *Econometrica* 48, 1805-1813.
- Griffiths, W.E. and P.A.A. Beesley (1984), The small sample properties of some preliminary test estimators in a linear model with autocorrelated errors, *Journal of Econometrics* 25, 49-61.
- Griffiths, W.E. and G.G. Judge (1990), Testing and estimating location vectors under heteroscedasticity, Mimeo, University of California, Berkeley.
- Gujarati, D.N. (1988), *Basic Econometrics*, 2nd ed., Singapore, McGraw-Hill.
- Gun, A.M. (1967), Use of a preliminary test in the estimation of factorial means, *Calcutta Statistical Association Bulletin* 16, 49-72.
- Gurland, J. and R.S. McCullough (1962), Testing equality of means after a preliminary test of equality of variances, *Biometrika* 49, 403-417.
- Harvey, A.C. (1976), Estimating regression models with multiplicative heteroscedasticity, *Econometrica* 44, 461-465.

- Harvey, A.C. (1990), *The Econometric Analysis of Time Series*, 2nd ed., London, Phillip Allan.
- Han, C-P., C.V. Rao and J. Ravichandran (1985), Inference based on conditional specification: A second bibliography, *Communications in Statistics: Theory and Methods* A 17, 1945-1964.
- Hausman, J. (1978), Specification tests in econometrics, *Econometrica* 46, 1251-1271.
- Hildreth, C. and J.P. Houck (1968), Some estimators for a linear model with random coefficients, *Journal of the American Statistical Association* 63, 584-595.
- Honda, Y. (1982), On tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal, *The Manchester School of Economic and Social Studies* 50, 116-125.
- Imhof, P. (1961), Computing the distribution of quadratic forms in normal random variables, *Biometrika* 48, 419-426.
- Jayatissa, W.A. (1977), Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal, *Econometrica* 45, 1291-1292.
- Judge, G.G. and M.E. Bock (1978), *The Statistical Implications of Pre-test and Stein Rule Estimators in Econometrics*, Amsterdam, North-Holland.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T-C. Lee (1985), *The Theory and Practice of Econometrics*, 2nd ed., New York, John Wiley and Sons.
- Judge, G.G. and M.E. Bock (1983), Biased estimation, in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Amsterdam, North-Holland, 599-649.
- Judge, G.G. and T.A. Yancey (1981), Sampling properties of an inequality restricted estimator, *Economics Letters* 7, 327-333.
- Judge, G.G. and T.A. Yancey (1986), *Improved Methods of Inference in Econometrics*, Amsterdam, North-Holland.
- King, M.L. (1979), Some aspects of statistical inference in the linear regression model, Ph.D. Thesis, University of Canterbury.
- King, M.L. (1981), The alternative Durbin Watson test: An assessment of Durbin and Watson's choice of test statistic, *Journal of Econometrics* 71, 51-66.
- King, M.L. (1983a), Testing for autoregressive against moving average errors in the linear regression model, *Journal of Econometrics* 21, 35-51.
- King, M.L. (1983b), Testing for moving average regression disturbances, *Australian Journal of Statistics* 25, 23-34.

- King, M.L. (1984), A new test for fourth-order autoregressive disturbances, *Journal of Econometrics* 24, 269-277.
- King, M.L. (1987), Testing for autocorrelation in linear regression models: A survey, in M.L. King and D.E.A. Giles (eds.) *Specification Analysis in the Linear Model*, London, Routledge and Kegan Paul, 19-73.
- King, M.L. (1989), Testing for fourth-order autocorrelation in regression disturbances when first order autocorrelation is present, *Journal of Econometrics* 41, 285-301.
- King, M.L. and D.E.A. Giles (1977), A note on Wallis' bounds test and negative autocorrelation, *Econometrica* 45, 1023-1026.
- King, M.L. and D.E.A. Giles (1984), Autocorrelation pre-testing in the linear model: Estimation, testing and prediction, *Journal of Econometrics*, 25, 35-48.
- King, M.L. and M. McAleer (1987), Further results in testing AR(1) against MA(1) disturbances in the linear regression model, *Review of Economic Studies* 54, 649-663.
- Kiviet, J.F. (1980), Effects of ARMA errors on tests for regression coefficients: Comments on Vinod's article; improved and additional results, *Journal of the American Statistical Association* 75, 353-358.
- Koerts, J. and A.P.J. Abrahamse (1969), *On the Theory and Application of the General Linear Model*, Rotterdam, Rotterdam University Press.
- Krämer, W. (1989), The robustness of the Chow test to autocorrelation among disturbances in P. Hackel (ed.) *Statistical Analysis and Forecasting of Structural Change*, Berlin, Springer-Verlag.
- Leonard, T. (1977), A Bayesian approach to some multinomial estimation and pre-testing problems, *Journal of the American Statistical Association* 72, 869-874.
- L'Esperance, W., D. Chall and D. Taylor (1976), An algorithm for determining the distribution function of the Durbin-Watson test statistics, *Econometrica* 44, 1325-1326.
- Lipsey, R.G. and J.M. Parkin (1970), Incomes policy: A re-appraisal, *Econometrica* 37, 115-138.
- Mandy, D.M. (1984), The moments of a pre-test estimator under possible heteroscedasticity, *Journal of Econometrics*, 25, 29-33.
- Marquardt, D.W. and R.D. Snee (1975), Ridge regression in practice, *The American Statistician* 29, 3-20.
- McCallum, B.T. (1972), Relative asymptotic bias from errors of omission and measurement, *Econometrica* 40, 757-758.
- Mills, E.S. (1962), *Price, Output and Inventory Policy*, New York, Wiley.

- Mittlehammer, R.C. (1984), Restricted least squares, pre-test, OLS and Stein rule estimators: Risk comparisons under model misspecification, *Journal of Econometrics* 25, 151-164.
- Morey, M. (1984), The statistical implications of preliminary specification error testing, *Journal of Econometrics* 25, 63-72.
- Nadiri, M.I. (1969), The determinants of real cash balances in the U.S. total manufacturing sector, *Quarterly Journal of Economics* 83, 173-196.
- Nakamura, A. and M. Nakamura (1978), On the impact of the tests for serial correlation upon the test of significance for the regression coefficient, *Journal of Econometrics* 7, 199-210.
- Newey, W.K. and K.D. West (1987), A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* 45, 203-208.
- Ohtani, K. (1983), Preliminary test predictor in the linear regression model including a proxy variable, *Journal of the Japan Statistical Society* 13, 11-19.
- Ohtani, K. (1985), A note on the use of a proxy variable in testing hypotheses, *Economics Letters* 17, 107-110.
- Ohtani, K. (1988), Optimal levels of significance of a pre-test in estimating the disturbance variance after the pre-test for a linear hypothesis on coefficients in a linear regression, *Economics Letters* 28, 151-156.
- Ohtani, K. and T. Toyoda (1978), Minimax regret critical values for a preliminary test in pooling variance, *Journal of the Japan Statistical Society* 8, 15-20.
- Ohtani, K. and T. Toyoda (1980), Estimation of regression coefficients after a preliminary test for homoscedasticity, *Journal of Econometrics* 12, 151-159.
- Ohtani, K. and T. Toyoda (1985a), Testing linear hypotheses on regression coefficients after a pre-test for disturbance variance, *Economics Letters* 17, 111-114.
- Ohtani, K. and T. Toyoda (1985b), Small sample properties of tests of equality between sets of coefficients in two linear regressions, *International Economic Review* 26, 37-44.
- Ohtani, K. and T. Toyoda (1988), Small sample properties of the two-stage test of equality of individual coefficients after a pre-test for homoscedasticity in two linear regressions, *Journal of the Japan Statistical Society* 18, 23-35.
- Ozcam, A. and G.G. Judge (1991), Some risk results for a two stage pre-test estimator in the case of possible heteroscedasticity, *Journal of Econometrics* 48, 355-371.

- Ozcam, A., G.G. Judge, A. Bera and T. Yancey (1991), The risk properties of a pre-test estimator for Zellner's seemingly unrelated regression model, Mimeo, Department of Agricultural and Resource Economics, University of California, Berkeley.
- Park, R.E. (1966), Estimation with heteroscedastic error terms, *Econometrica* 34, 888.
- Pesaran, M.H. (1973), Exact maximum likelihood estimation of a regression equation with a first-order moving-average error, *Review of Economic Studies* 40, 529-536.
- Prais, J.J. and J. Aitchison (1954), The grouping of observations in regression analysis, *Review of the International Statistical Institute* 22, 1-22.
- Sawa, T. and T. Hiromatsu (1973), Minimax regret significance points for a preliminary test in regression analysis, *Econometrica* 41, 1093-1106.
- Schmidt, P. and R. Sickles (1977), Some further evidence on the use of the Chow test under heteroskedasticity, *Econometrica* 45 1293-1298.
- Sclove, S.L., C. Morris and R. Radhakrishnan (1972), Non-optimality of preliminary-test estimators for the multinormal mean, *Annals of Mathematical Statistics* 43, 1481-1490.
- Searle, S.R. (1982) *Matrix Algebra Useful for Statistics*, New York, John Wiley and Sons.
- Small, J.P. (1991), The exact powers of some autocorrelation tests when the disturbances are heteroscedastic, Discussion Paper 9111, Department of Economics, University of Canterbury.
- Szroeter, J. (1978), A class of parametric tests for heteroscedasticity in linear econometric models, *Econometrica* 46, 1311-1328.
- Stein, C. (1964), Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, *Annals of the Institute Statistical Mathematics* 16, 155-160.
- Taylor, W.E. (1977), Small sample properties of a class of two stage Aitken estimators, *Econometrica* 45, 497-508.
- Taylor, W. (1978), The heteroscedastic linear model: Exact finite sample results, *Econometrica* 46, 663-675.
- Thomas, J.J. and K.F. Wallis (1971), Seasonal variation in regression analysis, *Journal of the Royal Statistical Society A* 134, 57-72.
- Toro-Vizcarrondo, C.E. (1968), Multicollinearity and the mean square error criterion in multiple regression: A test and some sequential comparisons, Ph.D. thesis, North Carolina State University.

- Toro-Vizcarrondo, C.E. and T.D. Wallace (1968), A test of the mean square criterion for restrictions in linear regression, *Journal of the American Statistical Association* 63, 558-572.
- Toyoda, T. (1974), Use of the Chow test under heteroscedasticity, *Econometrica* 42, 601-608.
- Toyoda, T. (1976), Optimal pre-testing procedure in prediction, *Kokumin Keizai-gaku Zasshi* 132, 72-92. (In Japanese.)
- Toyoda, T. and K. Ohtani (1986), Testing equality between sets of coefficients after a preliminary test for equality of disturbance variances in two linear regressions, *Journal of Econometrics* 31, 67-80.
- Toyoda, T. and T.D. Wallace (1975), Estimation of variance after a preliminary test of homogeneity and optimal levels of significance for the pre-test, *Journal of Econometrics* 3, 395-404.
- Toyoda, T. and T.D. Wallace (1976), Optimal critical values for pre-testing in regression, *Econometrica* 44, 365-375.
- Varian, H.R. (1975), A Bayesian approach to real estate assessment, in S.E. Fineberg and A. Zellner (eds.), *Studies in Bayesian Econometrics and Statistics in Honour of Leonard J. Savage*, Vol. I, Amsterdam, North Holland, 195-208.
- Vinod, H.D. (1976), Effects of ARMA errors on the significance tests for regression coefficients, *Journal of the American Statistical Association* 71, 929-933.
- Walker, A.M. (1967), Some tests for separate families of hypotheses in time series analysis, *Biometrika* 54, 39-68.
- Wallace, T.D. (1972), Weaker criteria and tests for linear restrictions in regression, *Econometrica* 40, 689-698.
- Wallace, T.D. (1977), Pre-test estimation in regression: A survey, *American Journal of Agricultural Economics* 59, 431-443.
- Wallace, T.D. and V.G. Ashar (1972), Sequential methods in model construction *Review of Economics and Statistics* 54, 172-178.
- Wallis, K.F. (1972), Testing for fourth order autocorrelation in quarterly regression equations, *Econometrica* 40, 617-636.
- Wan, A.T.K. (1992), The sampling performance of inequality restricted and pre-test estimators in mis-specified linear regressions, Discussion Paper 9207, Department of Economics, University of Canterbury.
- Watson, G.S. and E.J. Hannan (1956), Serial correlation in regression analysis II, *Biometrika* 43, 436-448.

- Watt, P.A. (1979), Tests of equality between sets of coefficients in linear regressions when disturbance variances are unequal: Some small sample properties, *The Manchester School of Economics and Social Studies* 47, 391-396.
- Weerahandi, S. (1987), Testing regression equality with unequal variances, *Econometrica* 55, 1211-1216.
- White, H. (1980), A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* 48, 817-838.
- White, K.J., S.D. Wong, D. Whistler and S.A. Haun (1990), *SHAZAM Econometrics Computer Package: User's Reference Manual, Version 6.2*, New York, McGraw-Hill.
- Wickens, M.R. (1972), A note on the use of proxy variables, *Econometrica* 40, 759-761.
- Wong, J.K. and J.A. Giles (1991), Optimal critical values of a preliminary test for linear restrictions in a regression model with multivariate Student-t disturbances, Discussion Paper 9114, Department of Economics, University of Canterbury.
- Yancey, T.A., G.G. Judge and S. Miyazaki (1984), Some improved estimators in the case of possible heteroscedasticity, *Journal of Econometrics* 25, 133-150.
- Yancey, T.A., G.G. Judge and R. Bohrer (1989), Sampling performance of some joint one-sided preliminary test estimators under squared error loss, *Econometrica* 57, 1221-1228.